A BISTABLE GENETIC SWITCH CONTROLS ANTIBIOTIC RESISTANCE TRANSFER IN ENTEROCOCCUS FAECALIS

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

ANUSHREE CHATTERJEE

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

ADVISOR: WEI-SHOU HU
CO-ADVISOR: YIANNIS N. KAZNESSIS

JULY, 2011

ACKNOWLEDGEMENTS

I am grateful to my Ph.D. advisor Wei-Shou Hu for being a constant source of guidance and support throughout my phD work. His attention to details and quest for perfection has driven me to become a better researcher, and for that, I am immensely thankful. His enthusiasm to follow new ideas and motivation to drive them through to success has inspired me over the past five years. Thank you for this great intellectual journey! I would like to thank Gary M. Dunny for his guidance and support. I will surely remember the scientific discussions we would have. I have learned a lot from those discussions. Your inputs during our collaboration have been invaluable. I would to thank my co-advisor Yiannis N. Kaznessis for his guidance and encouragement over the past five years. I am grateful to Prof. Doraiswami Ramkrishna at Purdue University for all the intriguing intellectual discussions and the hospitality when I was at Purdue. I am grateful to Prof. Eriko Takano and Dr. Sarika Mehra for their valuable inputs during my phD research.

A big thanks to past and present member of the Hu group, Dr. Anne Kantardjieff, Dr. Salim Charaniya, Dr. Marlene Castro, Siguang Sui, Bhanu C. Mulukutla, Nitya M. Jacob, Huong Le, Jason D. Owens, Shikha Sharma, Kathryn Johnson, Andrew Yonky, Nandita Vishwanathan, Tyler P. Price, Ravali Raju and Dong Seong Cho for providing a fun as well as a stimulating work environment. A special thanks to bunch of wonderful undergraduates that have worked with me, Laurie Drews, Hyacinth Cui and Michael Srienc. A very big thank you to Dunny lab members, Dawn Manias, Kristi L. Frank, Christopher M. Johnson, Laura C. Cook, Thinh Le, Frank, Nick,

Rachael, Aaron Barnes, Suzzane Grindle, Olivia Chuang, Katie Ballering and Heather Haeming. I have really enjoyed working at my second bench in Mayo building in the Dunny lab! Some special mentions, Chris, thanks for sharing your knowledge! Laura, thanks for always wishing me best whenever I would head down to New Mexico. Dawn, thanks for teaching me cloning and many other molecular biology techniques. Aaron Barnes for teaching me how to do microscopy on cells as tiny as *Enterococcus faecalis*, your willingness to help me with confocal microscopy is truly appreciated! A special thanks to Che-Chi Shu from Ramkrishna lab at Purdue University, thank you, for all the good work. A warm thank you to past and present Kaznessis group members, Dr. Howard Salis, Dr. Vassilios Sotiropoulos, Dr. Poonam Srivastava, Dr. Dan Bolintineanu, Katherine, Konstantinos, Victor, Emilia, Fatimeh and Patrick. I would like to acknowledge LeAnn Higgins at the Center for Mass Spectrometry and Proteomics for helping me with proteomics data.

In the end, I would like to thank my parents and my brother for their constant love and support during course of my life and for inspiring me to keep moving forward. Last but not least, I would like to thank my very dear husband, Prashant, for always being there for me. You inspire me to be better in every way I can, life is beautiful with you in it.

Dedication

To my husband, my parents and my brother

ABSTRACT

The recent rise in microbial drug resistance is a growing challenge for future therapy of bacterial infections. Increased prevalence of antibiotic resistance in bacteria is an outcome of evolution via natural selection. However, the built-in design feature of bacteria to transfer DNA containing antibiotic resistance both within the same species and across species is the main culprit for the spread of drug resistance. One of the main factors driving the rise of drug resistant microbes is the transfer of antibiotic resistance genes present on mobile plasmids between donor and recipient cells via the mechanism of conjugation. In order to combat microbial drug resistance, novel strategies need to be developed to block such transmission of antibiotic resistance. In this work, the gene regulatory components involved in transfer of tetracycline resistance confers plasmid pCF10 between plasmid-carrying donor cells and plasmid-deficient recipient cells in bacterium *Enterococcus faecalis* is investigated. In the native state the donor cell exists in an OFF or conjugation-incompetent state. A pheromone released by the recipient cells serves as the chemical trigger for switching the donor cell from OFF to an ON or conjugation-competent state. The onset of conjugation is tightly regulated via multi-layered regulation offered by two-key genes *prgQ* and *prgX* present on pCF10 in response to the pheromone secreted by recipient cells. Using mathematical modeling and experimentation, we describe a novel mechanism of gene-regulation due to transcriptional interference and sense-antisense RNA interaction as a result of convergent transcription in the *prgX/prgQ* operon. We demonstrate that such a multi-layered gene-regulatory mechanism confers the system a bistable genetic switch controlling

conjugative gene transfer between donor and recipient cells. A similar regulatory advantage offered by convergent transcription in attributing a bistable switch-like behavior in the *scbA-scbR* operon controlling antibiotic production in *S.coelicolor* is also investigated.

Both mathematical model and experiments demonstrate that donor cells also control response to pheromone by changing the number of copies of pCF10 plasmid inside the cell. Cells with higher copies show increased robustness of the bistable switch and lower sensitivity to pheromone. Once bistable genetic-switch is ON, expression of genes encoding various proteins involved in the transfer of the plasmid are induced, however, this also causes production of an inhibitor of conjugation, thus giving rise to negative feedback loop which causes the donor to return to OFF state. Modeling and experimental analysis of dynamic response to induction indicate that this negative feedback loop causes a brief surge of expression of the entire operon. We show that the inhibitor signaling peptide for pCF10 based system, acts as quorum-sensing signal with the role of turning-OFF conjugation at a population-wide scale. An interplay of positive and negative feedback loops allows the donor cell to quickly transition between ON and OFF states and is critical both for the transfer of plasmid and survival of the cell. Studying both the turning-ON and turning-OFF mechanisms of the switch allows identification of potential drug targets for blocking transmission of antibiotic resistance for use in future therapy.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 The Antibiotic Resistance Game

Antibiotics have been an extremely important weapon in the fight against infections for over half a century. However, increasing bacterial antibiotic resistance is a growing reality (CDC report (2008)). According to World Health Organization report, within a decade or two the current antibiotics-based therapies may become ineffective (WHO report, 2000). This is primarily due to two reasons. Firstly, bacteria and other microorganisms that are known to cause infection are remarkably resilient and can develop antibiotic resistance as an outcome of evolution via natural selection (Koirala and Thapa, 2011). When antibiotics are administered, bacteria containing the resistance survive and multiply to become the predominant population. Secondly and more importantly, resistant bacteria pass on their resistance to other bacteria through a mechanism called "conjugation". During conjugation, resistance-conferring genes typically present on a plasmid (a circular piece of DNA), are transferred from one bacterium to the other, thereby propagating drug resistance throughout the bacterial population (Figure 1.1). Plasmids often carry genetic information which enhances the chances of survival for the bacteria (Bergstrom *et al.*, 2000; Ghigo, 2001). The transfer of plasmids among bacteria is associated with bacterial adaptation and evolution, thus

giving rise to important phenomena such as enhanced bioremediation capabilities, gene flow from genetically modified microorganisms and potentially dangerous antibiotic resistance by pathogens (Gogarten and Townsend, 2005).

Over the last few decades a number of bacterial strains have become multi-drug resistant, including *staphylococci, enterococci, streptococci, salmonella, Mycobacterium tuberculosis* and others (Bergstrom *et al.*, 2000; Koirala and Thapa, 2011). Multidrug resistance is defined as the ability of a living cell to show resistance to variety of structurally and functionally unrelated drugs. Recently, the *Enterococcus faecalis* strain V583 has been reported to be resistant to vancomycin, an antibiotic which is has been traditionally reserved as a drug of last resort (Aakra *et al.*, 2005). *Enterococcus faecalis* has been known to transfer antibiotic resistance not only within species but also across species. The first step towards countering antibiotic resistance is to develop a fundamental understanding about the mechanisms of resistance transfer. It is important to note, that an effort to develop new antibiotics would be futile if the organism would be resistant to them in due course of time (often referred to as "the resistance game").

This thesis is an investigation of bacterial cell-cell communication controlling the transfer of clinically isolated tetracycline resistance plasmid pCF10 in bacterium *Enterococcus faecalis* (*E.faecalis*) via conjugation. *E. faecalis* normally resides in the human intestine as a component of the natural flora and in recent years has emerged as 3rd most common hospital acquired infection (Tendolkar *et al.*, 2003). In the past few years, the genetic components involved in the regulation of conjugation have been identified by

our collaborator Prof. Dunny at the Department of Microbiology (Dunny *et al.*, 1978; Dunny, 2007a; Dunny and Johnson, 2011). The bacterial population is divided into two sub-populations, the donor cells, which carry the antibiotic resistant plasmid, and recipient cells, which lack the plasmid and are antibiotic sensitive. Potential recipient cells release a peptide pheromone (a 7 amino acid signaling molecule called cCF10) which is sensed by donor cells (Figure 1.1) (Buttaro *et al.*, 2000; Mori *et al.*, 1988). Once, a donor senses the proximity of recipient cells, a certain set of genes are expressed within the donor, which allows formation of a mating channel between the donor and recipient cell, with the transfer of a copy of the plasmid into the recipient (Chen *et al.*, 2008). Once the recipient obtains this plasmid, it comes a transconjugant (i.e. a recipient that now behaves as a donor), thus further propagating resistance within the population (Dunny, 2007).

To develop a strategy to stop resistance transfer, it is crucial to understand which mechanisms allow the donor cells to switch between a "conjugationally-competent" or "ON" state, vs. a "conjugationally –incompetent" or "OFF" state. In this work, we investigate such a switch behavior using a combined mathematical and experimental approach. A fresh look from a perspective of systems analysis greatly enhances our understanding of this system. To summarize, this thesis is an engineering perspective to a biologically relevant question.

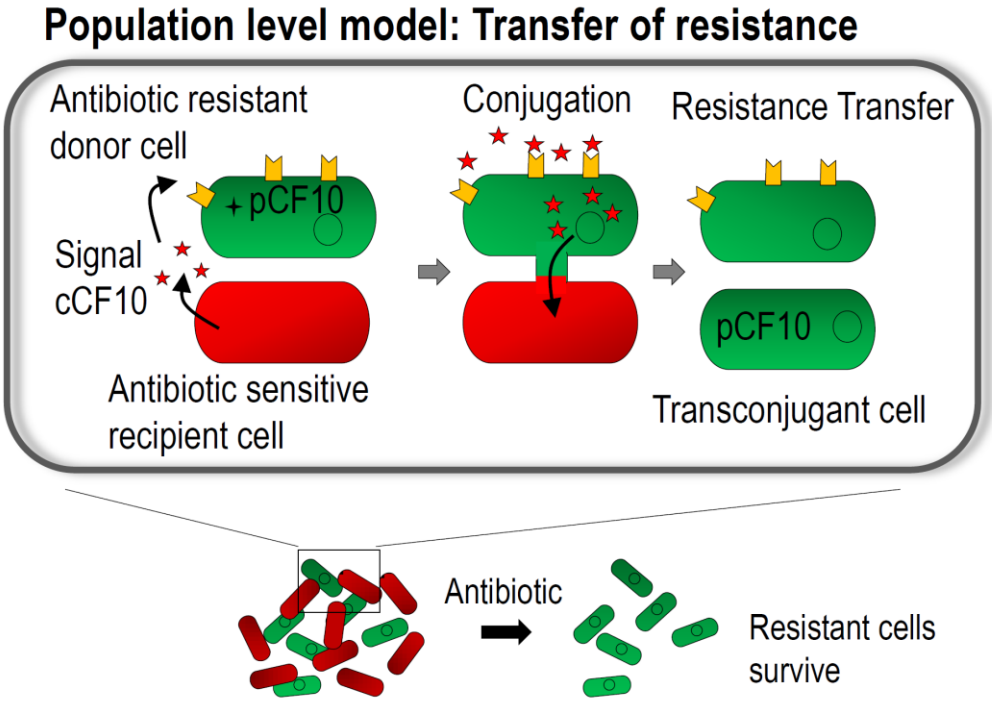


Figure 1.1: The antibiotic resistance game. Transfer of antibiotic resistance between donor and recipient cells.

1.2 Scope of the Thesis

In this thesis, we investigate pheromone-induced transfer of antibiotic resistance plasmid pCF10 in *Enterococcus faecalis* via conjugation between donor cells and recipient cells. Gene regulatory components of the genetic switch controlling the onset of conjugation are studied using both mathematical and experimental approaches to identify the key regulatory elements and perturb genetic switch controlling conjugation as a strategy to counter drug resistance transfer (Chapter 4 and 5). A key finding of the current research includes studying the biological function of convergent (or antisense)

transcription with focus on the regulatory role of RNA polymerase traffic (Transcriptional Interference) and antisense RNA in biological switch behavior (Chatterjee *et al.*, 2011). The role of convergent transcription in conferring a bistable switch controlling transfer of antibiotic resistance conferring plasmid pCF10 in bacterium *Enterococcus faecalis* is investigated. Mathematical models based on a reduced genetic-network developed for the pCF10 based system were employed to gain insights into the regulatory landscape of the pCF10 conjugation system (Chapter 5). Mathematical model guided experimental design led to discovery of a novel mechanism in the control of antibiotic resistance transfer this plasmid (Chapter 4). The mathematical model developed for pCF10 was modified to investigate the role of convergent transcription in Butyrolactone regulon in *Streptomyces coelicolor* in conferring a bistable genetic switch for antibiotic synthesis in Chapter 8 (Chatterjee *et al.*, 2011).

The model on the pCF10 system is further extended to include the effect of pCF10 plasmid copy number on bistable switch behavior in this system and consequently the efficiency of antibiotic resistance transfer in *E. faecalis* (Chapter 6) (Cook *et al.*, 2011). Model prediction and experimental data indicate that population heterogeneity caused by varying pCF10 copy number in donor cells can adjust their response to induction, such that cells with higher pCF10 copy number require a high level of induction to turn ON conjugation.

Along these lines, we elucidate the mechanism by which pCF10 conjugation system allows an induced donor cell to turn-OFF conjugation (Chapter 7) (Chatterjee *et*

al., (to be submitted)). We show that the inhibitor signaling peptide for pCF10 based system acts as a quorum-sensing signal with the role of turning-OFF conjugation at a population-wide level. Unlike conventional quorum sensing systems, where the signaling molecule turns-ON gene expression at a population scale, on indicate that this negative feed-back loop causes a brief "pulse-like" expression of the entire operon. Interplay of positive and negative feedback loops allows the system to quickly transition between ON and OFF states and reduces the metabolic burden on the donor cell.

In the past few years the genetic elements involved in the regulation of this intriguing signaling system have been identified. Much progress has been made and a mechanism consistent with experimental observations has been proposed. The emphasis of the research has been on the molecular structure of the regulatory proteins and their interactions with the target elements. With the complexity and the non-linear nature of the system, a fresh look at this system from a perspective of systems analysis greatly enhances our understanding of its regulation.

1.3 Thesis organization

This thesis is arranged in nine chapters. Chapter 2 reviews the current knowledge about the genetic regulation involved in transfer of plasmid pCF10 in *E. faecalis* and introduces basic concepts about transcriptional interference and antisense RNA regulation, which would be alluded to frequently in this thesis. Chapter 3 includes a brief review of concept of bistable genetic switch in biological reaction networks. Chapter 4 describes the mechanisms which regulate the genetic switch controlling conjugation of

plasmid pCF10 in *E. faecalis,* one of the key findings of this thesis. Chapter 5 outlines the details of the mathematical model used in Chapter 4 and later in Chapter 6 and 7. In chapter 6, the role of pCF10 plasmid copy number on the response of donor cells to pheromone causing transfer of resistance is discussed, especially its role in resistance dissemination under native growth conditions such as biofilm cells. In chapter 7, a newly discovered role of the previously known inhibitor of conjugation, as a quorum sensing signaling molecule used by donor cells to sense their own concentration is discussed. Chapter 8 highlights the regulatory advantage offered by convergent transcription in attributing a bistable switch-like behavior in the *scbA-scbR* operon controlling antibiotic production in *S. coelicolor*. Finally, Chapter 9 summarizes the key findings and conclusions of this thesis and sheds light on some of the future directions of this research.

CHAPTER 2

BACKGROUND

2.1 Enterococcus faecalis: A Primer

Enterococcus faecalis is a gram positive bacterium that is a member of the normal intestinal microflora of humans and other animals. It is also found in wide variety of ecological niches associated with either animal or plant host or free-living in nature. Although, in healthy humans this bacterium is not pathogenic, it can cause serious extra-intestinal infections in individuals with impaired immune systems (McCormick *et al.*, 2002). *E. faecalis* is one of the leading causes of nosocomial bacteremia, surgical wound infection, endocartidis and urinary tract infection (Carniol and Gilmore, 2004; Koch *et al.*, 2004; McCormick *et al.*, 2002) and is one the three most frequently hospital-acquired infections and a major concern for the health care system (Tendolkar *et al.*, 2003). *E. faecalis* possesses two unique properties that accounts for its ecology and medical importance. Firstly, *E. faecalis* is an extremely hardy organism that can adapt its physiology to survive in harsh conditions, for example it possesses a very hard cell wall which makes this bacterium resistant to killing under many conditions lethal for common microbes, including starvation, desiccation, and presence of antimicrobial compounds

such as detergents, disinfectants and antibiotics. Secondly, *E. faecalis* is very proficient in acquisition of genetic mobile elements encoding traits such as virulence and antibiotic resistance transfer (Harbottle *et al.*, 2006). This property of *E. faecalis* to develop antibiotic resistance makes it resistant to many and in some cases all standard therapies, thereby contributing to its medical relevance. Recently, analysis of genome sequence of *E. faecalis* strain V583, the only strain of the species whose complete genome sequence has been reported, indicated that nearly 30% of the genome represents mobile genetic elements acquired by lateral gene transfer from other organisms (Paulsen *et al.*, 2003). Exogenously acquired genome elements has conferred *E. faecalis* with resistance to the strongest antibiotic known, vancomycin (VRE) and in some cases all known antibiotics (Aakra *et al.*, 2005).

An effective control of multiple drug resistant entercocci will require a better understanding of physiology, genetics, ecology and pathogenic potential of these organisms. The pheromone induced plasmid transfer in *E. faecalis* results in transmission of drug resistance in the bacterial population. This phenomenon resembles quorum sensing based cell to cell communication system, a method used widely in bacterial systems.

2.2 Quorum sensing: cell to cell communication in bacteria

Quorum sensing (QS) is the ability of cells to sense their cell density, communicate with each other and behave as a population instead of a single cell. QS involves production, release, detection and response to chemical signals that increase in

concentration as a function of cell density. Cells release a chemical signal into the environment (Fuqua *et al.*, 1996; Fuqua *et al.*, 1994). Once the signal concentration reaches a threshold level, it leads to induction of target genes in the cells. In some cases, the target genes include those needed for signal production, thereby amplifying the response; hence the term autoinducer (AI) is often used to designate the signal molecules in these systems. The target genes regulated in this fashion often encode activities that when expressed in large number of cells would benefit the entire population, but might not work effectively at extremely low levels when produced by a single cell as illustrated in Figure 2.1. The concentration of autoinducer in a nanomolar range or below is often sufficient for cells to respond to the chemical signal and thus participate in biological activity (Dunny and Leonard, 1997).

Bacteria use QS communication circuits to regulate diverse array of physiological activities such as virulence, antibiotic production, conjugation, sporulation, biofilm formation etc. QS is observed in large number of bacterial species, including, *Vibrio fischeri, Pseudomonas aeruginosa, Agrobacterium tumefaciens, Streptococcus pneumonia, Bacillus subtilis* and *Enterococcus faecalis* (Miller and Bassler, 2001; Waters and Bassler, 2005). Gram negative bacteria such as *E. coli* have been known to use acylated homoserine lactones (AHL) as autoinducers, on the other hand "low G+C" gram positive bacteria such as *E. faecalis* have been known to use processed oligo-peptides as autoinducers and "high G+C" gram positive bacteria such as *Streptomyces coelicolor* have been known to use γ-butyrolactone signal molecules for communication (Hughes

and Sperandio, 2008; Takano, 2006; Waters and Bassler, 2005){Kleerebezem, 1997 #460}. Every QS system has its own unique chemical signal and the cognate gene network. This uniqueness cause bacteria to communicate effectively with one another, coordinate gene expression and therefore control the behavior of the entire community.

In gram negative bacteria chemical signals generally diffuse though the cell wall (Miller and Bassler, 2001; Waters and Bassler, 2005). In case of gram positive bacteria using signaling peptides, it is believed that many of the peptides function by binding to membrane associated receptors of the Histidine kinase family. However, *Enterococcus faecalis* employs a different kind of peptide signal reception system. In *E. faecalis* the peptide is imported into the cell via an ATP-dependent transport system, where it binds to a target receptor membrane protein which allows import of the peptide into the cell. In *E. faecalis* the process of conjugation is induced by a QS mechanism. The recipient cells release the chemical signal which is intercepted by donor cells, which get induced and coordinately express genes required for conjugation once the recipient cells reach a threshold concentration (Dunny and Leonard, 1997). In the following section we will look into the mechanism of plasmid transfer via conjugation.

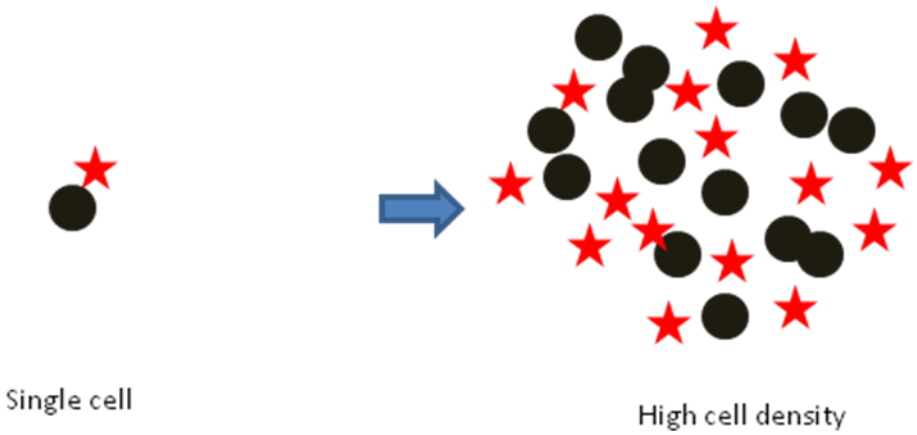


Figure 2.1: Quorum sensing. Cell (●) respond once the chemical signal (★) has reached a threshold density which depends on the cell density.

2.3 Conjugation: a mechanism of gene transfer in Entercocci

A very common phenomenon observed in bacteria is the transfer of genetic material via conjugation. Discovered in 1946 by Joshua Lederberg and Edward Tatum (Lederberg and Tatum, 1946), bacterial conjugation is a mechanism of horizontal gene transfer which involves physical cell-to-cell contact. As illustrated in Figure 2.2, Horizontal gene transfer (HGT) or lateral gene transfer (LGT), is any process in which an organism transfers genetic material to a cell which is not its offspring, in contrast vertical gene transfer (VGT) is a process where an organism transfers genetic material to its offspring (Gogarten and Townsend, 2005; Jain *et al.*, 1999; Kurland *et al.*, 2003). Horizontal gene transfer is very common among bacteria, even between different bacterial species. This process is thought to be significant in increased drug resistance,

since when one bacterium acquires resistance, it can quickly transfer the resistance genes to other bacteria, both in the same and across different species.

There are three ways of horizontal gene transfer, transformation (Chen and Dubnau, 2004), transduction (Young, 1993) and conjugation (Sorensen *et al.*, 2005). Transformation refers to uptake of free DNA by a bacterial cell and its stable integration into the bacterial genome. Transduction involves incorporation of foreign circular DNA, formed by accidental packaging of cellular DNA into bacteriophage particles during replication, into the host bacterial genome. Conjugation is a highly specific process whereby DNA is transferred from a donor to recipient bacteria by a specialized multi-protein complex, called the conjugation complex (Grohmann *et al.*, 2003). In gram negative bacteria, this physical contact between cells is established by complex extracellular filaments called sex pili (Grohmann *et al.*, 2003). In gram positive bacteria, the means to achieve the cell to cell contact have not been established so far, but it is thought that induced donor cells express conjugation related genes which cause donor and recipient cells to adhere to each other allowing formation of a physical mating channel though which the plasmid is transferred into the recipient cell (Figure 2.2).

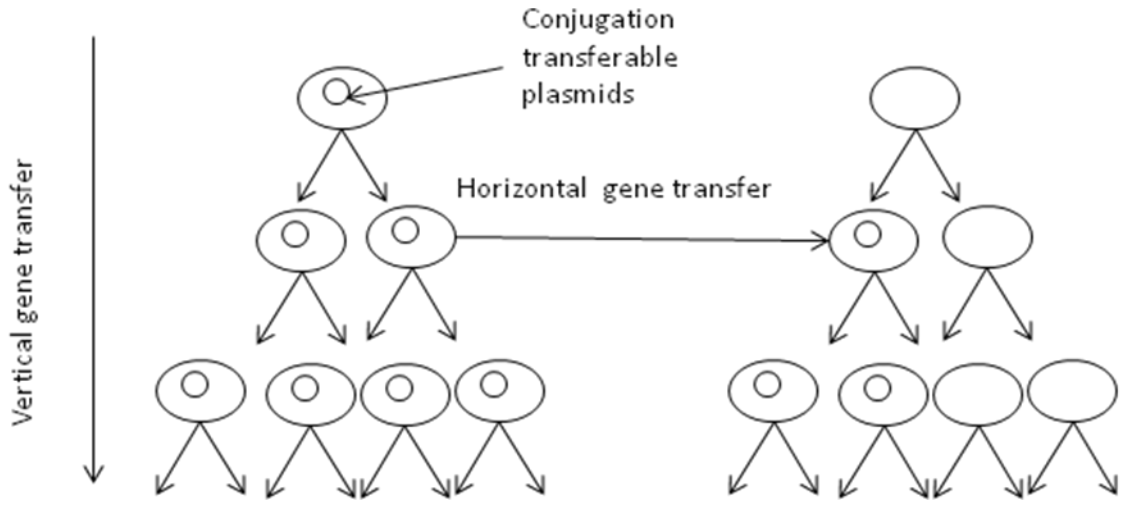


Figure 2.2: A schematic view of vertical and horizontal modes of plasmid transfer. A copy of plasmid can be transferred to from the donor cell to a recipient cell (plasmid-free) and be stably inherited in the daughter cells of the recipient (Simonsen, 1991).

E. faecalis is known to acquire mobile genetic elements via conjugation (Clewell *et al.*, 2002; Dunny *et al.*, 1982). The most efficient form of enterococcal conjugation involves transfer of mobile elements called pheromone-inducible conjugative plasmids from one cell to the other. In general, quorum sensing systems synchronize a single cell type population, where the cells releasing the chemical signals are the ones with altered gene expression leading to a synchronized population behavior. The complexity and unique features of *E. faecalis* pheromone system reside in synchronizing two kinds of cell types in the bacterial population, the donor cells and the recipient cells. The donor cells contain the antibiotic resistance plasmid in them, whereas the recipient cells are plasmid deficient. In the *E. faecalis* system the recipient cell releases the chemical signal into the environment (Chandler and Dunny, 2004). The chemical signal is encoded by a gene on

the chromosome of the recipient cell. The genetic elements responsible for sensing and responding to the chemical signal are present on the conjugation transferable plasmid present in the donor cells. As in the case of many bacteria, these plasmids carry a set of transfer genes that encode proteins which mediate attachment of donor cells to recipient cells, formation of a mating channel connecting the cytoplasmic compartments of the two cells and enzymatic processing of the plasmid DNA such that both the donor and recipient cells have at least a single copy of the plasmid. Upon transfer of the plasmid to the recipient cell, it attains the drug resistance gene on the plasmid; it is now called a transconjugant (Figure 2.3). The transconjugant can become itself a donor of the plasmid and participate in the process of conjugation. The process of conjugation is an energetically costly process for the donor cell. The donor cell needs to be sure that there are enough recipient cells in its surrounding environment before becomes conjugation competent.

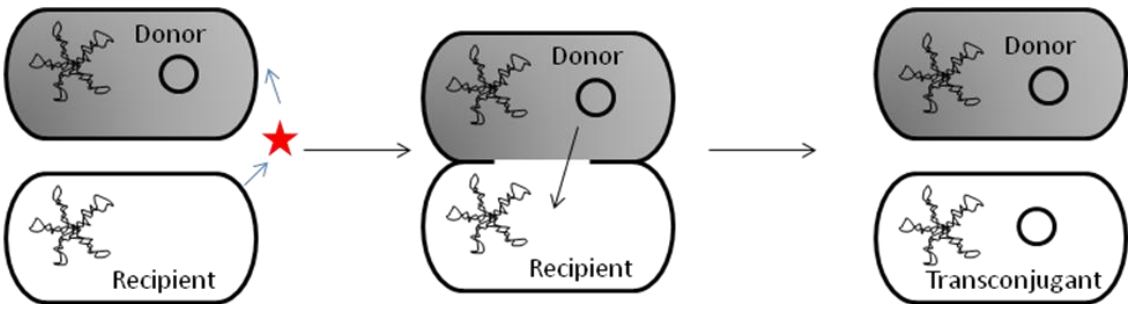


Figure 2.3: The process of conjugation

The genetic makeup of the plasmid entails a mechanism to prevent the donor cell from responding to false signals. In the cell this is achieved by the presence of robust genetic networks which make the process of conjugation a highly regulated process. As

the process of conjugation proceeds, the recipient cell population gradually becomes "identical" to the donor cell. This way antibiotic resistance plasmid is propagated in the population. However, conjugation ceases before the entire recipient population is converted to donor cells. The next section reviews the genetic regulation involved in the process of conjugation for *E. faecalis*.

2.4 Genetic regulation in Enterococcus faecalis during conjugation

Three pheromone plasmids have been studied extensively for the enterococcal system: pCF10, pAD1 and pPD1 (Chandler and Dunny, 2004). We are interested in the pheromone plasmid pCF10. The plasmid pCF10 is a 58 kb conjugative plasmid which was isolated from *E. faecalis* clinical strain SF-7 (Dunny *et al.*, 1981). The tetracycline resistance was conferred to the plasmid by stable incorporation of conjugative transposon Tn925 (Carniol and Gilmore, 2004; Torres *et al.*, 1991) (Figure 2.4). There are several other families of conjugative plasmid that have been identified in *E. faecalis* clinical isolates (Chandler and Dunny, 2004). In each of these systems, the plasmid-free recipient cells produce several pheromones, each specific to a cognate plasmid (Dunny *et al.*, 2001; Mori *et al.*, 1988). The donor cells respond to the pheromone and get induced to participate in conjugation. In the pCF10 system, there are two opposing factors that regulate conjugation. Along with the inducer pheromone released by recipient cells, there is an inhibitor molecule released by the donor cells that prevents conjugation. Depending on the relative concentrations of the inducer and the inhibitor molecules in the

environment conjugation is regulated. In the following the genetic composition of pCF10 plasmid is described.

2.4.1 Conjugation transferable plasmid pCF10

The sequence analysis of conjugative plasmid pCF10 was reported by Hirt et. al., Figure 2.4 shows the plasmid map for pCF10 (Hirt *et al.*, 2005). The plasmid consists of 67,673 bp with the tetracycline resistance transposon T925 accounting for 18,032 bp. The plasmid can be divided into three sections depending on the function of genes; these are regulatory loci devoted to pheromone sensing, formation of mating aggregates and DNA processing machinery along with the origin of transfer (*oriT*) target sequence. Genes present on each of these loci participate in the process of conjugation. It has been shown that the even though different parts of the plasmid are coordinately regulated; these parts seem to have evolved from three different sources (Hirt *et al.*, 2005). There are regions in the pCF10 that share homology with other pheromone responsive plasmids pAD1 and pPD1, suggesting that the pheromone-sensing system originated from a common ancestor. The nomenclature for pCF10 open reading frames (ORF) is shown in Table 2.1.

The region between *prgN-prgQ* encodes for pheromone sensing and control, plasmid replication and maintenance. We will mainly focus on this region. The pheromone responsive genes (*prg*), respond to the pheromone cCF10 and inhibitor iCF10 molecules and hence regulate conjugation. The region from *prgZ-prgQ* is involved in detection of pheromone and initiation of donor response to exogenous pheromone signal. This genetic loci is also responsible for preventing self-induction. The segment from

prgQ-pcfH is coordinately expressed when exogenous pheromone cCF10 is added, with the expression of the remaining genes on pCF10 and chromosome being unaffected (Hirt *et al.*, 2005).

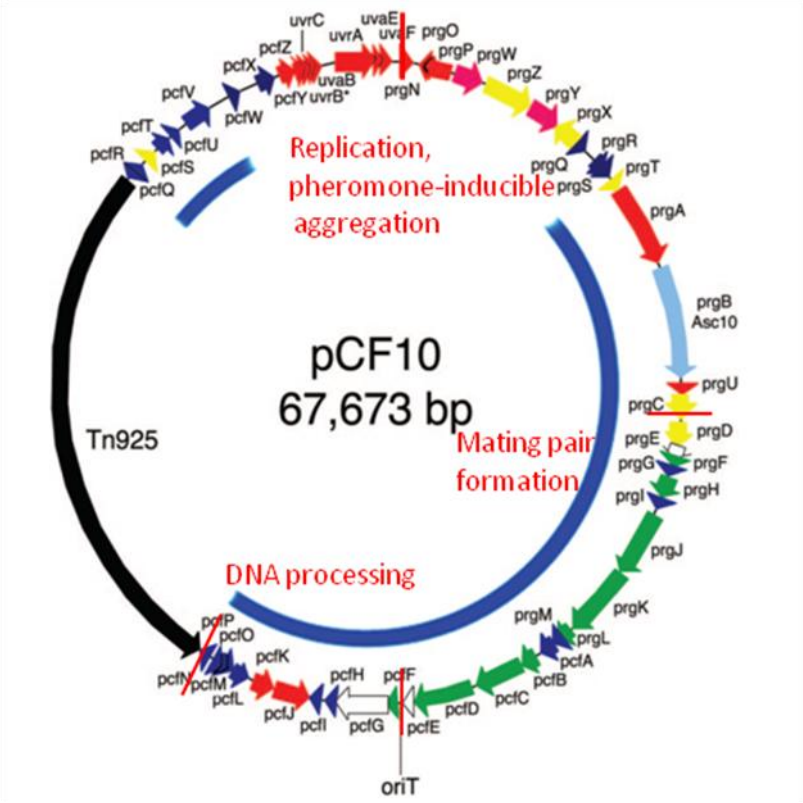


Figure 2.4: The genes encoded on plasmid pCF10 (Hirt *et al.*, 2005).

Table 2.1 List of genes on plasmid pCF10

ORF	Gene	pCF10 location	Orientation	Gene/protein size (bp/AA)	Homology	Identity/similarity (%)	Function
1	*uvrA*	66603–258	cw	1,392/443	*uvrA* (pAD1)	99	UV resistance
2	*uvaE*	255–605	cw	351/116	*orfB* (pAD1)	96/96	
3	*uvaF*	562–774	cw	213/70	*orfC* (pAD1)	94/97	
4	*prgN*	1207–1503	cw	297/98	*orfE* (pAD1)	91/97	
5	*prgO*	1935–1606	ccw	330/109	*orf57* (pRE25)	81/90	
6	*prgP*	2788–1859	ccw	930/309	*orf58* (pRE25)	79/90	
7	*prgW*	3035–4036	cw	1,002/333	*repA* (pPD1)	95/98	Replication
8	*prgZ*	4195–5832	cw	1,638/545	*traC* (pAM373)	38/59	Pheromone receptor
9	*prgY*	5843–6997	cw	1,155/384	*traB* (pPD1)	77/91	Prevention of self-induction
10	*prgX*	7983–7030	ccw	954/317	*traA* (pAM373)	25/45	Negative regulator
11	*prgQ*	8192–8263	cw	72/23			iCF10
12	*prgR*	8721–9125	cw	405/134			
13	*prgS*	9125–9397	cw	273/90			
14	*prgT*	9699–9887	cw	189/62	EP0043 (pAM373)	69/85	
15	*prgA**	10015–12690	cw	2,677/891	*sea 1* (pAD1)	82/90	Surface exclusion
16	*prgB**	12883–16800	cw	3,918/1,305	EF0005 (PI, *E. faecalis*)	96/97	AS
17	*prgU**	16862–17218	cw	357/118	*orf3* (pAD1)	98/100	
18	*prgC**	17246–18103	cw	859/285	*EP0037* (pAM373)	63/74	
19	*prgD**	18145–19044	cw	900/299	*EP0038* (pAM373)	25/47	
20	*prgE**	19064–19498	cw	435/144	BK5–T (*L. lactis*)	37/58	ssb
21	*prgF**	19545–19772	cw	228/75	*gbs1142*	26/54	
22	*prgG**	19786–20082	cw	297/98	EFB0016 (pTEF2)	100	
23	*prgH**	20097–20897	cw	801/206	*gbs1362*	33/55	
					gbs140	29/52	
24	*prgI**	20899–21252	cw	354/117	EFB0018 (pTEF2)	100	
25	*prgJ**	21359–23596	cw	2,238/745	*gbs1135*	41/64	
26	*prgK**	23608–26223	cw	2,616/871	*gbs1359*	35/50	
					gbs1133	31/47	
27	*prgL**	26247–26873	cw	627/208	*gbs1132*	41/60 (129 aa)	
28	*prgM**	26860–27105	cw	246/81	EFB0022 (pTEF2)	98/100	
29	*pcfA**	27089–27697	cw	609/202	EFB0023 (pTEF2)	97/98	
30	*pcfB**	27856–28341	cw	487/161	*gbs1129*	27/55	
31	*pcfC**	28341–30170	cw	1,830/609	*gbs1128*	44/64	TrsK-like
32	*pcfD**	30221–32380	cw	2,160/719	*gbs1126*	34/54	
					ltrC (pMRC01)	31/49	
33	*pcfE**	32413–32685	cw	273/90	*ltrD* (pRS01)	36/56	
34	*pcfF**	32931–33287	cw	357/118	*gbs*1122	30/61	
35	*pcfG**	33288–34973	cw	1,686/561	*ltrB* (pRS01)	50/68	Relaxase
					gbs1121	38/61	
36	*pcfH**	35006–35356	cw	351/116	EFB0031 (pTEF2)	100 (72 aa)	
37	*pcfI**	35680–35847	cw	168/55	EFB0032 (pTEF2)	100	
38	*pcfJ**	37210–35873	ccw	1,338/445	ORF63 (pAD1)	81/90	
39	*pcfK**	37983–37207	ccw	777/258	ORF62 (pAD1)	91/94	
40	*pcfL**	38674–38126	ccw	549/182	EFB0035 (pTEF2)	100	
41	*pcfM**	38889–38698	ccw	192/63	EFB0036 (pTEF2)	90/90	
42	*pcfN**	39059–38874	ccw	186/61	EFB0037 (TEF2)	100	
43	*pcfO**	39514–39065	ccw	450/149	EFB0038 (pTEF2)	100/100	
	pcfO′	39619–39701	ccw	84/28	orf60 (pAD1)	67**	
		39881–39075		195/65			
44	*pcfP**	39801–40001	cw	201/66	EFB0040 (pTEF2)	100	
45	*pcfQ**	58372–58142	ccw	231/76	EFB0041 (pTEF2)	100	
46	*pcfR**	58427–58921	cw	495/164	EFB0042 (pTEF2)	99/100	
47	*pcfS**	58987–59439	cw	453/150	EP0029 (pAM373)	65/76	ssb
48	*pcfT**	59617–60195	cw	579/192	EF0045 (pTEF2)	91/91	Thermonuclease
49	*pcfU**	60277–60606	cw	330/109	EFB0046 (pTEF2)	97/100	
50	*pcfV**	60829–61854	cw	1,026/341	EFB0047 (pTEF2)	100 (pTEF2: 486 aa)	
51	*pcfW*	62776–63048	cw	273/90			
52	*pcfX*	63710–64369	cw	660/219	BT2225 *Bacteroides thetaiotaomicron*	31/54	
53	*pcfY*	64451–65071	cw	621/206	orf86 (pAD1)	95/98	DNA invertase
54	*pcfZ*	65079–65375	cw	297/98	EFA0074 (pTEF1)	96/97	
55	*uvrC*	65369–65593	cw	225/74	*uvrC* (pAD1)	95/98	
56	*uvrB*	65654–65863	cw	183/60	*uvrB* (pAD1)	94/100**	
57	*uvaB*	65875–66048	cw	174/57	EFA0076 (pTEF1)	94/100 (truncated)	

The *prgA*, *prgB* (Bensing and Dunny, 1993) and *prgC* genes encode surface proteins that allow the donor and recipient cells to adhere to each other, thereby forming the mating channel. In between *prgQ* and *prgA* there are genes that seem to affect the expression of downstream genes after pheromone induction. The genetic loci between *prgC* and *pcfD* is largely unknown, but is suspected to participate in mating channel formation (Staddon *et al.*, 2004). The region between *pcfE-pcfH* encodes for DNA processing machinery with the region between *pcfE* and *pcfF* encodes for origin of transfer *oriT*. The genes in plasmid pCF10 are expressed in the order of events occurring in conjugation.

2.4.2 Release of mating pheromone cCF10 by recipient cell

The expression of a large set of gene products required for the transfer of plasmid pCF10 via conjugation is controlled by cell-cell communication between plasmid-free recipient cells and plasmid-carrying donor cells using a heptapeptide mating pheromone called cCF10 (Antiporta and Dunny, 2002; Mori *et al.*, 1988), released by the recipient cells (Figure 2.5). It has been shown that seven to eight amino acid peptides constitute the mature pheromones, which are processed from pre-cursor lipoproteins. Other known peptide pheromones known for *E .faecalis* cAD1, cPD1 are also processed from lipoproteins encoded within the chromosome (An and Clewell, 2002; Antiporta and Dunny, 2002). In *E. faecalis* both the recipient and donor cells contain a gene *ccfA* present on the chromosome which encodes a 22 amino acid lipoprotein called pre-cCF10, which forms the precursor for the mating pheromone cCF10 (Antiporta and Dunny, 2002;

Clewell *et al.*, 2000). An and Clewell (An and Clewell, 2002) identified a putative intra-membrane protease called Eep which is involved in the proteolytic cleavage of the pre-cCF10 lipoprotein with an amino acid sequence VKKYKRLLLMAGLVTLVFVLSACGTAPVS into the 7 amino acid peptide cCF10 with the sequence LVTLVFV (Figure 2.5). An and Clewell showed that Eep is responsible for the production of normal levels of several pheromones. They suggested that since the 7-8 amino acid mature pheromones peptides were more hydrophobic than the peptide precursors, the proteolysis may have been followed by active excretion of the pheromones into the environment. Following the processing in the membrane, Buttaro et. al. and Leonard et. al. showed that cCF10 can be found both in the growth medium as well as associated with the cell wall of recipient cells (Buttaro *et al.*, 2000, Leonard et al., 1996). After the processing in the membrane of the recipient cells, mature cCF10 is released into the culture medium. Once cCF10 is released into the environment, the cCF10 levels keep increasing in the environment, when a threshold concentration is reached donor cells are induced. The initiation of pheromone response in donor cells requires the binding of cCF10 to a membrane protein PrgZ, which is present on the cell wall of the donor cells and is encoded by the pCF10 plasmid. This is followed by import into the cytoplasm by a chromosomal oligopeptide permease ABC transporter. The scheme of events is shown in Figure 2.5. Since the *ccfA* gene is also present on the chromosome of the donor cell, the pCF10 plasmid has devised a mechanism that prevents

the donor cell to get induced to endogenously expressed cCF10. The next section describes the mechanism by which the donor cell prevents self-induction.

2.4.3 Donor cells in uninduced state: preventing self induction

The process of conjugation is an energetically costly process for the cell. As a result, the pCF10 plasmid in the donor cells has evolved a gene network which prevents the cell from getting induced under false conditions, such as self induction of a newly formed transconjugant. The *ccfA* and *Eep* genes involved in pheromone production are present on the chromosome; the donor cells also encode cCF10 molecules by the same mechanism as recipient cells. However, the donor cells prevent self-induction with the help of two plasmid encoded proteins, protein PrgY and an inhibitor molecule known as iCF10 (Chandler *et al.*, 2005).

The plasmid pCF10 (as shown in Figure 2.4) has a gene called *PrgY* (Chandler and Dunny, 2008; Leonard *et al.*, 1996) encoding a membrane protein PrgY which reduces the level of nascent cCF10 released by the donor cells by sequestering, modifying or degrading them. It has been shown that a major portion (about 60%) of prgY lies outside the cell membrane and is anchored by C-terminal transmembrane segments. The extracellular domain is known to be the functional domain of the protein as PrgY mutants with mutations in the N-terminal domain have been shown to lose function (Chandler *et al.*, 2005). Chandler et. al. suggested that PrgY has no interaction with the exogenously added pheromones.

PrgY is not sufficient for control of endogenous pheromone cCF10 levels in the donor cell. The donor cells possess significant levels of cCF10 in both the cell wall fractions and the culture medium to cause induction of donor cells (Buttaro *et al.*, 2000). The residual cCF10 in the donor cells is neutralized by an inhibitor molecule iCF10 encoded from the *prgQ* locus on the pCF10 plasmid (Buttaro *et al.*, 2000; Nakayama *et al.*, 1994). Both uninduced and induced state donor cells express a basal level of inhibitor molecule. In this way for the donor cell to be induced there has to be a high concentration of inducer cCF10, which happens only when the recipient cell density reaches a particular threshold density.

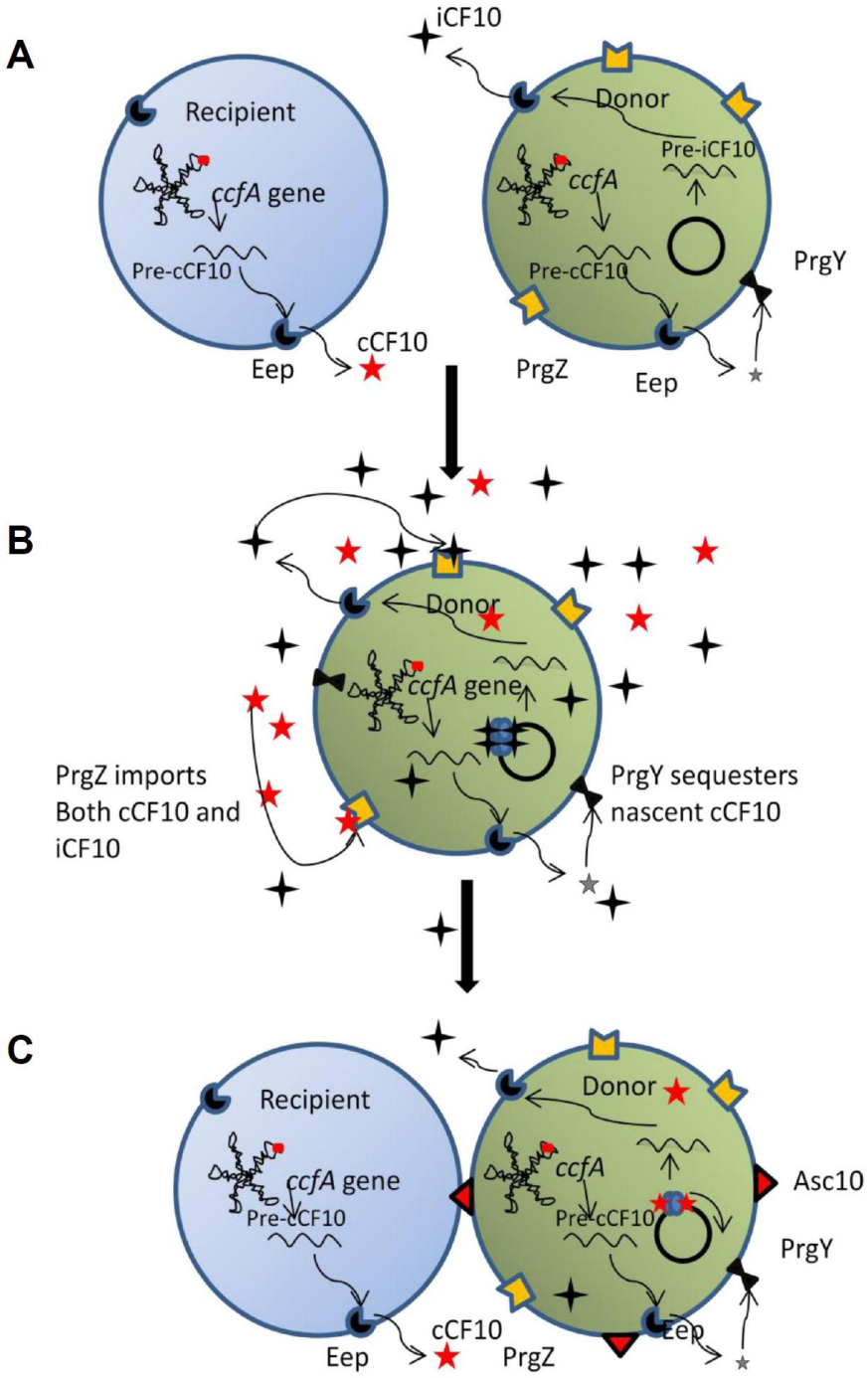


Figure 2.5: Scheme of events occurring during conjugation (adapted from (Shi *et al.*, 2005)). **(A)** Recipient cells (blue) and donor cells (green) release the inducer and inhibitor molecules respectively, **(B)** the donor cell gets induced, **(C)** the recipient and donor cells form a mating pair.

As described in greater detail in the next section, in the uninduced state the donor cells constitutively express a short transcript from the *prgQ* locus called Q_s which encodes a 22 amino acid pre-iCF10 peptide which is processed by the membrane protein Eep to give a heptapeptide inhibitor molecule iCF10 (Bae *et al.*, 2004; Bensing *et al.*, 1996; Chung *et al.*, 1995).The last codons of the pre-iCF10 encode the mature iCF10 sequence (AITLIFI). As in the case of cCF10, the processing of the mature peptide occurs simultaneously with peptide export. As a result iCF10 molecule exits the cell, and needs to be re-imported into the cell via binding with PrgZ membrane protein, to participate in regulation. The iCF10 and cCF10 levels in the culture medium have been reported to be in a delicate balance. HPLC fractionation of culture supernatants for *E. faecalis* strains containing pCF10 plasmid have shown that the ratio of iCF10 to cCF10 is about 50-100/L of culture volume (Hirt *et al.*, 2002; Nakayama *et al.*, 1994). At this ratio iCF10 just neutralizes cCF10 activity such that the donor cell is sensitive to extremely low concentration (10^{-11}M) of exogenous cCF10 that would be produced by a recipient cell in close proximity (Mori *et al.*, 1988).

2.4.4 The molecular switch: PrgX

PrgX is the primary cytoplasmic protein responsible for the negative regulation of pheromone-inducible conjugation functions of the plasmid pCF10. In the uninduced state the PrgX molecule is bound by the inhibitor molecule, which represses the transcription of downstream conjugation genes. In the induced state the PrgX molecule is bound by the inducer molecule which acts as a transcriptional activator for expression of downstream

conjugation genes. Both the inducer and inhibitor compete for binding to the PrgX molecule. This fine balance ensures that the donor cell is induced only under high cCF10 concentration which occurs at high recipient concentrations. In the following we look at this fine regulation in greater detail.

Crystal structures of the PrgX molecule reported by Shi et. al. have shown that PrgX is almost all helical, composed of 17 α-helices connected by loops of various lengths (Figure 2.6 (a)) (Shi *et al.,* 2005). It was shown that PrgX molecule has three function domains, an N-terminal domain containing the DNA binding domain, a central domain required for dimerization of PrgX molecule, and a C-terminal domain where the pheromones cCF10 and iCF10 bind (figure 2.6(b)) (Kozlowicz *et al.,* 2004; Shi *et al.,* 2005). It has been established that PrgX mutants are partially defective in DNA binding and oligomerization which affects the *prgQ* repression (Bae *et al.,* 2004; Kozlowicz *et al.*, 2004). Bae et. al. reported that PrgX has two DNA binding sites located in the *prgX-prgQ* intergenic region (Bae *et al.,* 2002). The primary binding site (operator O_1) which exhibits stronger affinity for PrgX consists of a 11 bp palindromic sequence (GTATTGAATAC) and is located 80 bases upstream of PrgX start site.

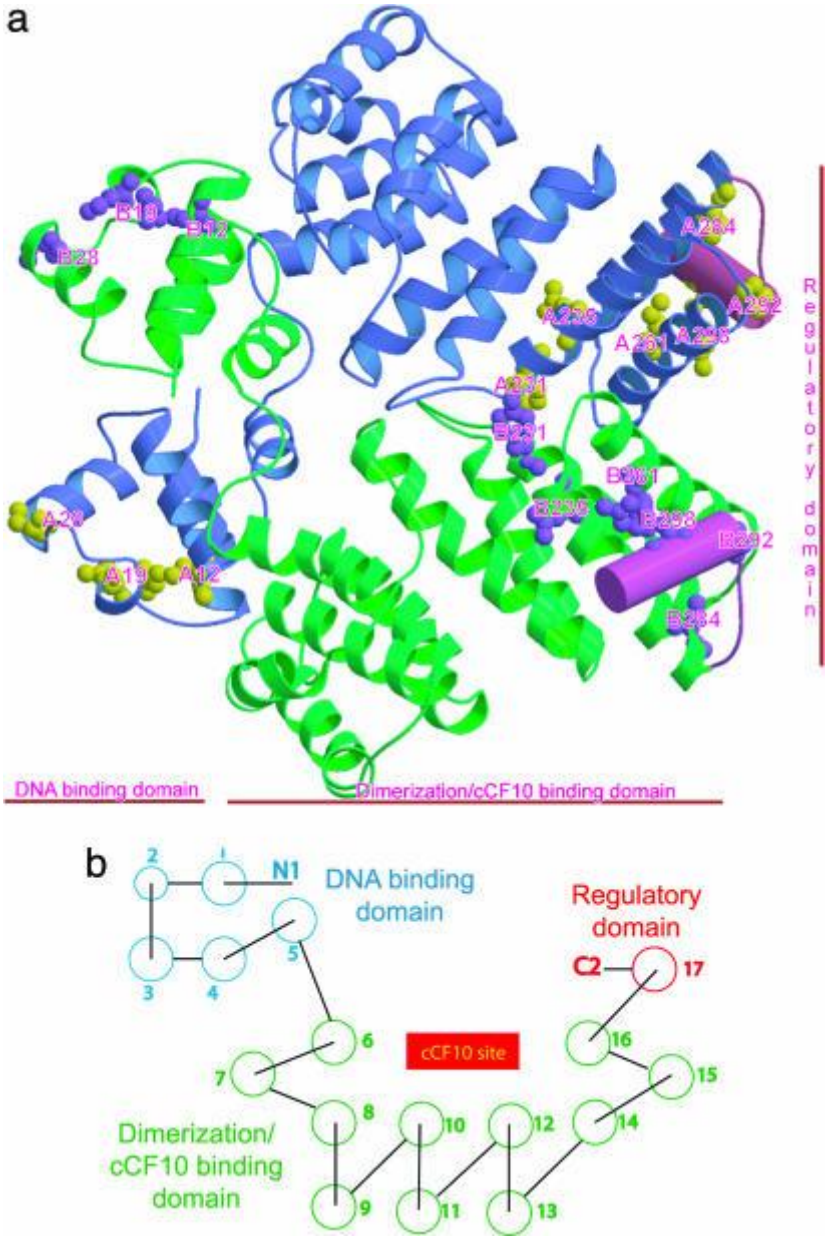


Figure 2.6: PrgX crystal structure. (a) A ribbon diagram for the PrgX homodimer showing the three functional domains of PrgX **(b)** The 17 α-helices are denoted by circles. The C-regulatory domain is shown in red, the DNA binding domain in blue and the dimerization domain in green, the cCF10 binding pocket is formed by two layers of α-helices (Shi *et al.*, 2005).

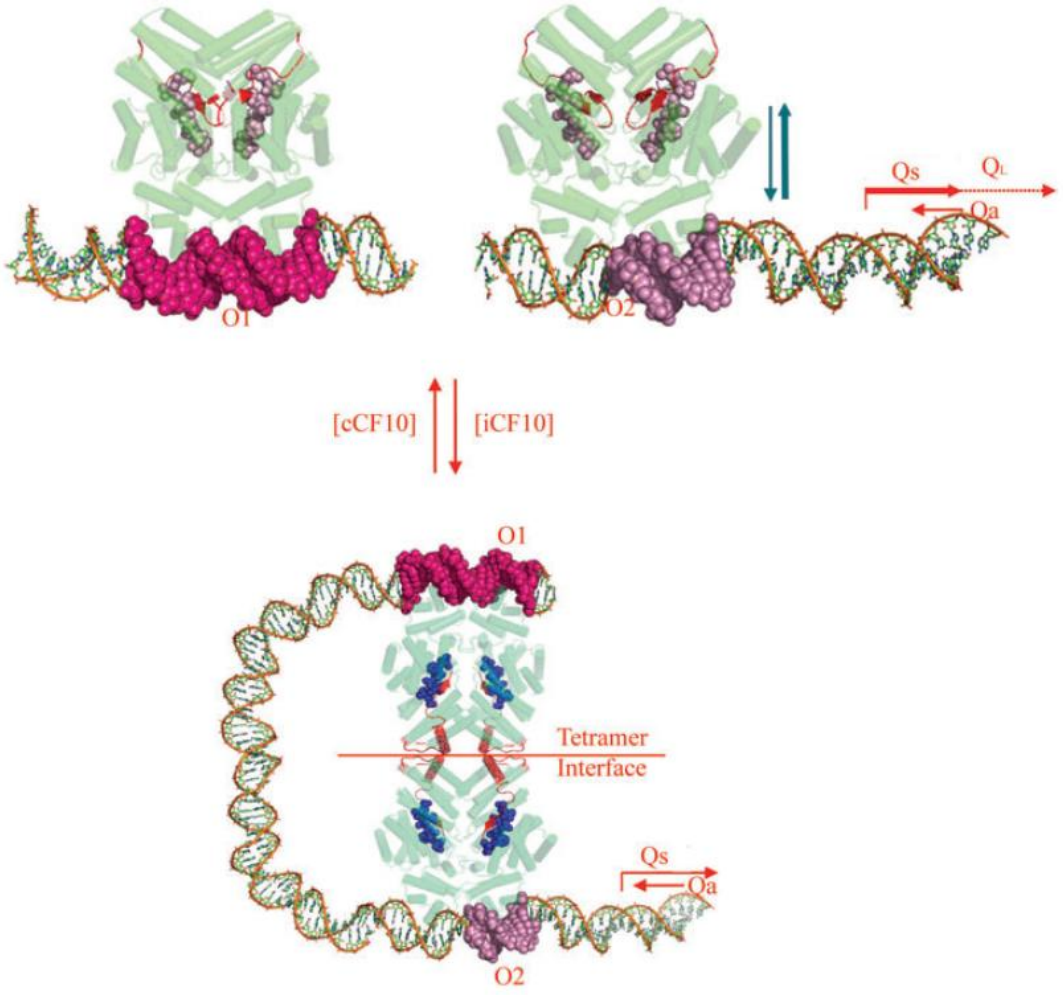


Figure 2.7: PrgX regulation of the *prgQ* operon. The PrgX molecule forms a tetramer conformation when bound to inhibitor iCF10 in the uninduced state causing the DNA to loop. Binding of cCF10 causes the tetramer conformation to break to give two dimers bound respectively to the O_1 and O_2 sites. Both iCF10 and cCF10 compete for binding to PrgX (Kozlowicz *et al.*, 2004).

A secondary binding site (operator O_2) is located between -35 and -10 regions of the PrgQ promoter (P_Q), which contains half of the palindromic sequence (GTATTGAATAC) and thus exhibits lower affinity for PrgX. Structural data have shown

that PrgX protein exists in an oligomeric state in the cell, binding of inducer molecule to the PrgX reduces oligomerization. Structural data show that the DNA binding sites O_1 and O_2 are bound by a PrgX dimer (Bae and Dunny, 2001; Bae *et al.*, 2002). The lower affinity Operator O_2 site overlaps the *prgQ* promoter such that occupancy of this site by PrgX is predicted to impede the ability of RNA polymerase to initiate transcription. The occupancy of the O_1 site greatly enhances the occupancy of the O_2 site, because of cooperativity.

Each PrgX molecule has a pocket where cCF10 and iCF10 can bind (Figure 2.6). The pocket is formed by several parallel and anti-parallel helices. Both cCF10 and iCF10 complete for binding to the PrgX molecule. In the uninduced cell, where the iCF10 out competes cCF10, the iCF10 binds to the PrgX molecule. In the induced state, cCF10 binds to the PrgX molecule (Shi *et al.*, 2005). PrgX dimers are stabilized by the interaction between the numerous residues of central segment of the protein, as well as domain swapping of N-terminal domains of adjacent monomers. In the uninduced state when the iCF10 molecule binds to PrgX, one set of $(PrgX\text{-}iCF10)_2$ dimer sits at the O_1 site, another $(PrgX\text{-}iCF10)_2$ dimer sits at the O_2 site (Figure 2.7).

When iCF10 binds to PrgX it promotes interaction between the two dimers to form a tetramer conformation. The iCF10 interacts with certain residues at the C-terminus of PrgX resulting in stabilization of tetramer conformation. This causes the DNA to loop (Figure 2.7). The looping of the DNA prevents the transcription of downstream conjugation genes. It is believed that the looped complex affects the

elongation, processing and decay of transcripts synthesized from the PrgX promoter by acting as a road block to RNA polymerase molecules.

When cCF10 occupies the binding pocket, a major rotation of approximately 20 amino acid residues occurs in the C-terminus. This causes the residues 296-298 to cover the bound pheromone in the pocket (Shi *et al.*, 2005), which causes a shift of conformation from a short α-helix to a β-sheet thereby moving the C-terminus. This results in the disruption of the tetramer structure. Disruption of the tetramer conformation causes a return to the dimer conformation and a break in the DNA looping. The un-looped DNA has a more open conformation with a reduced occupancy of the operator sites. The open conformation prevents blocking of RNA polymerase thereby leading to increased activity of the *prgQ* promoter. This mechanism is shown in Figure 2.7.

2.4.5 Transcriptional regulation of conjugation in pCF10

The switch from uninduced to induced state is a highly regulated process which is determined by the relative levels of the inducer molecule cCF10 and the inhibitor molecule iCF10 in the environment (Bensing *et al.*, 1996). Like many other cell-cell communication systems, the pCF10 *E. faecalis* system has a regulatory protein which acts as a transcriptional activator that regulates the expression of conjugation genes. In *E. faecalis* the pCF10 system this regulatory protein is PrgX. The regulation is achieved by the competitive binding of both the pheromones cCF10 and iCF10 to PrgX as explained in previous section. This section describes the post-transcriptional regulation occurring in pCF10 system. As seen in Figure 2.4 the plasmid map for pCF10 shows the *prgX* and

prgQ locus next to each other. In the following we describe the mechanism by which post-transcriptional regulation occurs. In uninduced cells the *prgQ* promoter gives rise to shorter transcripts which are only 400 nucleotide (nt) in length. These transcripts are referred to as Q_S (Q-short) (Bensing *et al.*, 1997). The 3' end of the Q_S transcript ends in the vicinity of an inverted repeat sequence (IRS1) which functions in termination of transcription as shown in Figure 2.7 (Bae *et al.*, 2004).

Induction by pheromone gives rise to increase in transcription from the *prgQ* promoter. This causes an increase in the extension of *prgQ* transcripts past the putative termination region IRS1 to give a transcript which is longer than the Q_S transcript by 130 nt and is termed as Q_L (Q-long). In the induced cells this 530 nt Q_L transcripts leads to the expression of downstream conjugation genes (Bensing and Dunny, 1993). The *prgX* and *prgQ* genes are adjacent to each other, but transcribed from opposite strands. The promoter for *prgX* lies in the *prgQ* operon, but on the non-template strand and is denoted as P_X (Figure 2.8). An unprocessed *prgX* mRNA contains a 130 nt non-coding sequence at its 5' end termed Anti-Q. Both the *prgQ* and *prgX* transcripts are subject to extensive processing. The *prgX* transcript is processed to give the *prgX* mRNA and Anti-Q transcript. The Anti-Q transcript is complimentary to a segment of the Q_S transcript and is known to sequester some Q_S via antisense mechanism. Genetic evidence has shown that Anti-Q prevents Q_s transcripts to extend beyond IRS1. This is because both Anti-Q and Q_S form extensive secondary structures, and the pairing of these two causes the 3' end of Q_S to behave as a functional terminator. Figure 2.9 demonstrates the change in

secondary structure of Qs from the anti-terminator to the terminator. In the uninduced cell the Q_S and Anti-Q transcripts exist in a balance such that none of the Q_s transcripts go beyond IRS1. However, when the cells are induced, the transcription of Q_S increases 2-5 times, this causes Q_S to titrate all the Anti-Q, the remaining Q_S can then go beyond IRS1 to form Q_L transcript. Genetic studies suggest that the Q_L RNA may activate the expression of *prgB* by increasing the processivity of transcription into the *prgB* region, thereby enhancing the transcription of *prgB*. The *prgB* gene expresses the surface protein Asc10 which then causes the donor and recipient cells to adhere and form the mating channel.

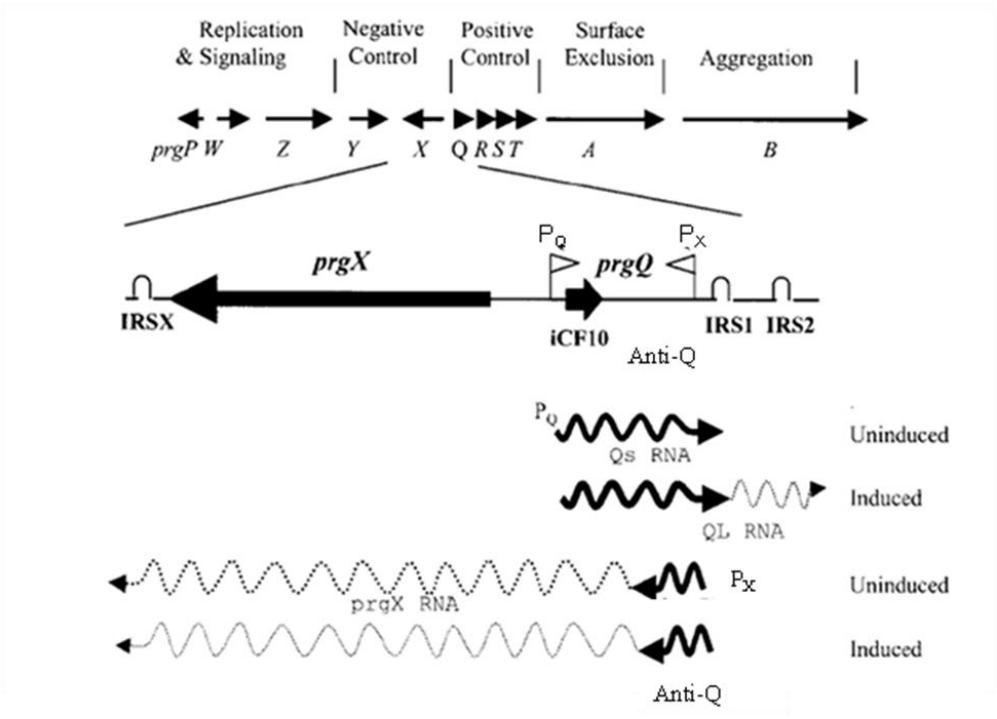


Figure 2.8: The genetic map of the pheromone responsive genes. The direction of transcription of genes is shown by bold arrows. The bottom part shows a detailed map for *prgX-Q* intergenic region. The promoters (*prgQ* promoter indicated as P_Q and *prgX* as P_X) are indicated by flags pointing in the direction of transcription. The inverted repeat sequence (IRSX, IRS1 and IRS2) is shown by inverted U's. The transcripts are shown under different induction state of the donor cell (Bae *et al.*, 2002).

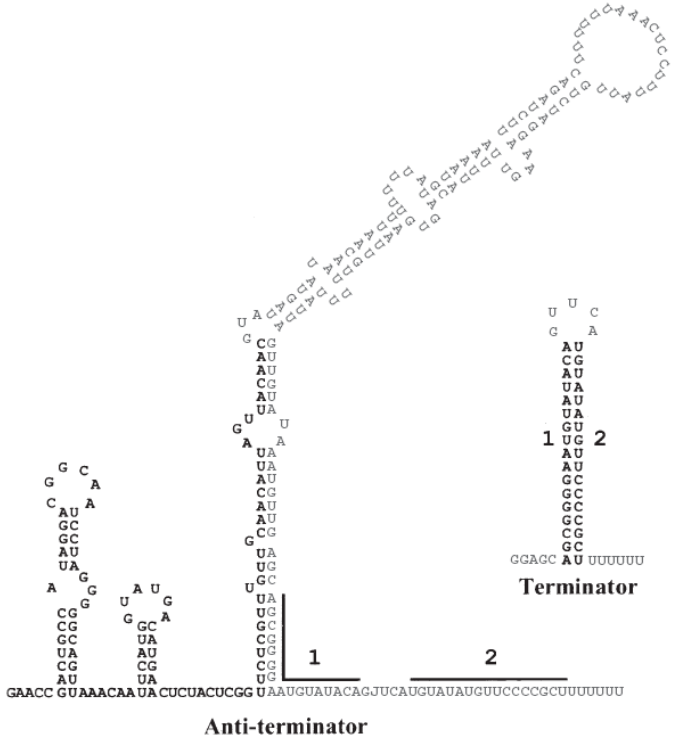


Figure 2.9: A model for Anti-QRNA mediated transcriptional termination. The secondary structure for Q_s RNA is shown. In presence of the Anti-Q terminator Q_S transforms from the anti-terminator structure to the terminator structure (Bae *et al.*, 2002).

2.4.6 Expression of downstream conjugation genes and plasmid transfer

Once the donor cell has been induced the downstream genes are transcribed. There is sufficient data to suggest that the longer transcript Q_L acts post-transcriptionally on the polycistronic message to enhance of translation of downstream gene *prgB* (Bensing *et al.*, 1997; Olmsted *et al.*, 1993). Expression of pCF10 *prgB* gene depends on the induction of donor cell. Only when the donor cell is induced is this gene expressed in significant amounts. The *PrgB* gene encodes a surface aggregation protein called Asc10

(Bensing and Dunny, 1997), which appears on the outer cell wall of the donor cell. This protein binds to the cell wall of a neighboring recipient cell causing donor-recipient cell aggregation. This allows for transfer of the plasmid from the donor to the recipient cell at high frequency.

How the Q_L transcript activates the transcription of *prgB* is not understood very clearly. Bae and other authors have (Bensing and Dunny, 1997; Chung *et al.*, 1995; Chung and Dunny, 1992) proposed a mechanism by which Q_L transcript could be involved in *prgB* gene expression. Bensing et. al. (Bensing and Dunny, 1997) predicted (Figure 2.10 (a)) the secondary structure of the Q_L transcript, and showed that the portion of the Q_L transcript interacts with the ribosome or other translational factors. It was shown that the 5' UTR of the *prgB* transcript was also able to interact with ribosomes. It has been proposed that there is likely interaction between Q_L and the helix 26a resulting in a conformational rearrangement of ribosomal RNA which could result in the release of an unidentified 16S rRNA domain that could interact with the *prgB* 5'UTR and hence support *prgB* translation (Bensing *et al.*, 1997). As shown in Figure 2.10 (b), in the uninduced state, due to formation of shorter transcripts (Q_S) due to transcriptional termination by RNA polymerase prevents the activation of prgB transcription. Figure 2.10 (c) shows the induced state. Bensing et.al. suggest that induction may cause a conformational change in the 3'UTR of *prgS* transcript, which allows the polymerase to become termination resistant. *PrgS* is also known to stabilize the association of Q_L with

the ribosomal components, an altered ribosome population results in enhanced *prgB* translation.

Upon sensing the peptide pheromone, *E. faecalis* transfers the plasmid pCF10 through a type IV secretion system (T4S). Once the surface protein is expressed other downstream conjugation genes are expressed. Recently Chen et. al. (Chen *et al.*, 2008) reported that an accessory factor *pcfF* and *pcfG* relaxase initiate the transfer of plasmid to the recipient cell by catalyzing a strand specific nick at the origin of transfer *(oriT)* (Figure 2.4). It is known that *pcfG* gene which encodes a relaxase protein. The relaxosome or the processed transfer intermediate interacts with a substrate receptor, PcfC. The mechanism by which this occurs is not yet clear. It has been observed that that expression of PcfG is spatially coordinated with the coupling protein PcfC. For pCF10, pcfC is a Type 4 coupling protein (T4CP) which has a N-terminal putative transmembrane domain with conserved NTP binding motifs, and other motifs that are common to the coupling protein family. This protein is responsible for delivering the nicked DNA substrate to a cognate transport apparatus for transfer of the DNA strand across the bacterial cell envelope into the recipient cell. In pCF10, the region between *prgB* and *pcfC*, an approximately 15 kb region, is predicted to encode for 13 mating pair formation (Mpf) proteins. These proteins are responsible for attachment of donor and recipient cells and the formation of tight mating junctions, with the formation of a membrane spanning channel between the donor and the recipient cells. Once the DNA substrate is transferred to the recipient cell, the plasmid is replicated using the rolling

circle mechanism of replication. In this manner the recipient cell also gets a copy of the plasmid pCF10, thereby achieving antibiotic resistance.

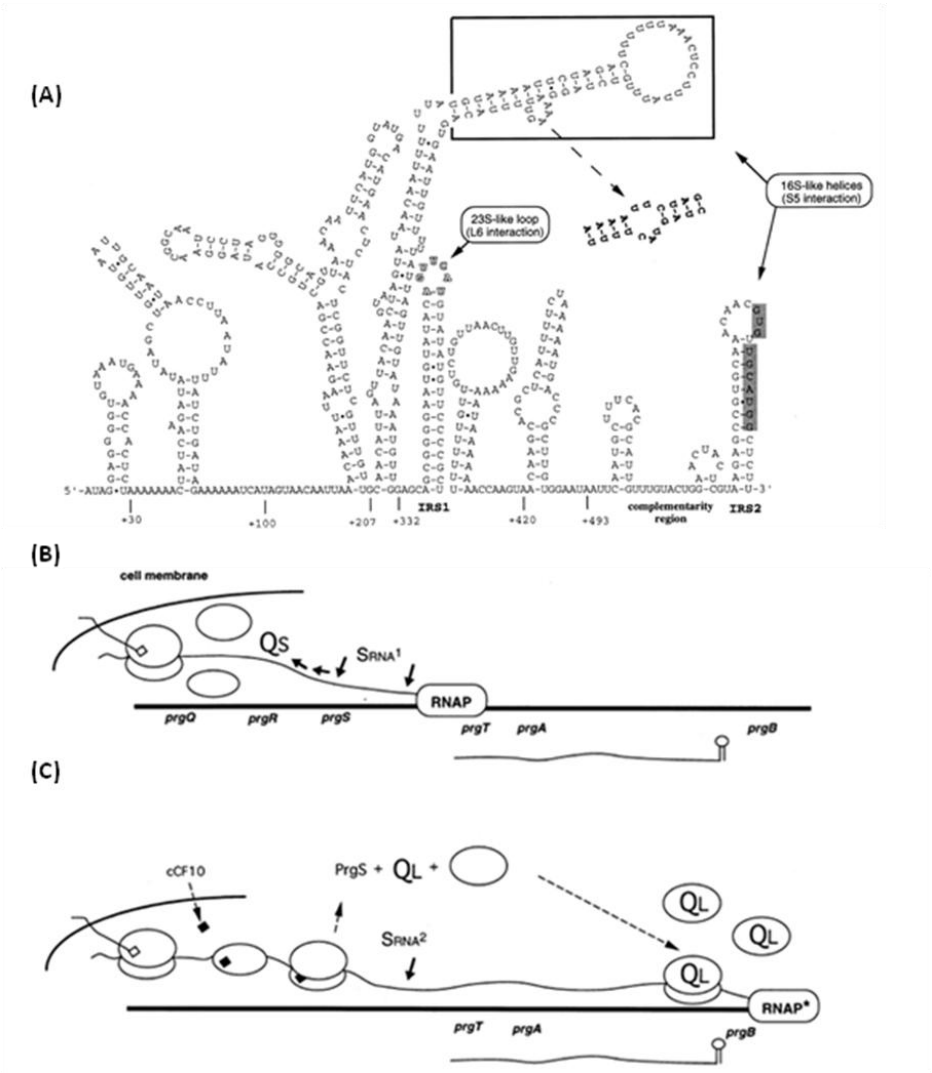


Figure 2.10: The model of activation of *prgB* transcripts. (A) Predicted secondary structure of QL with the domains of interaction shown **(B)** The model in the uninduced state **(C)** Induced state (see text for details) {Bensing, 1997 #398}.

2.6 Antisense Transcription: a widespread occurrence in genomes

Antisense transcription is a widespread phenomenon, present in numerous genes. A pair of genes are said to be antisense to each other, when these are present on opposite strands of DNA (one on sense and other on the antisense strand), with corresponding promoters convergent to each other, such that there is a partial overlap between transcripts (Figure 2.11 A). Such complementary transcripts are also known as *cis*-encoded Sense-Antisense transcripts (SAT). Recently, SAT's have been reported in viruses, prokaryotes, and eukaryotes such as yeast, Drosophila, Arabidopsis and mammals. Such SAT pairs are a result of convergent transcription. Convergent transcription has been speculated to represent a hidden layer of regulation (Hongay et al., 2006; Katayama et al., 2005).

The occurrence of SATs appears to be widespread within eukaryotic genomes; such as 5425 SAT's in mouse (Katayama et al., 2005); 2667 in human (Rodrigo Yelin et al., 2003); 1027 in *Drosophila melanogaster* (Misra et al., 2002), and 7600 in *Arabidopsis thaliana* (Yamada et al., 2003). Recent whole-transcriptome analyses of prokaryotic genomes have shown high levels of *cis*-antisense RNA, such as 3-13% in *L. monocytogenes* (Toledo-Arana et al., 2009), *M. pneumoniae* (Guell et al., 2009), *B. subtilis* (Liu et al., 2009), *Vibrio cholera* (Rasmussen et al., 2009), E.coli (Tjaden et. al., 2002). Experimental analysis has shown negative regulation of mRNAs through SAT interaction (Georg et al., 2009; Toledo-Arana et al., 2009); however, the mechanisms of regulation remain elusive in both prokaryotes and eukaryotes.

Two main mechanisms have been reported to operate among such sense and antisense transcripts, namely, transcriptional interference and antisense RNA interaction. In the following section, these mechanisms are described in greater detail.

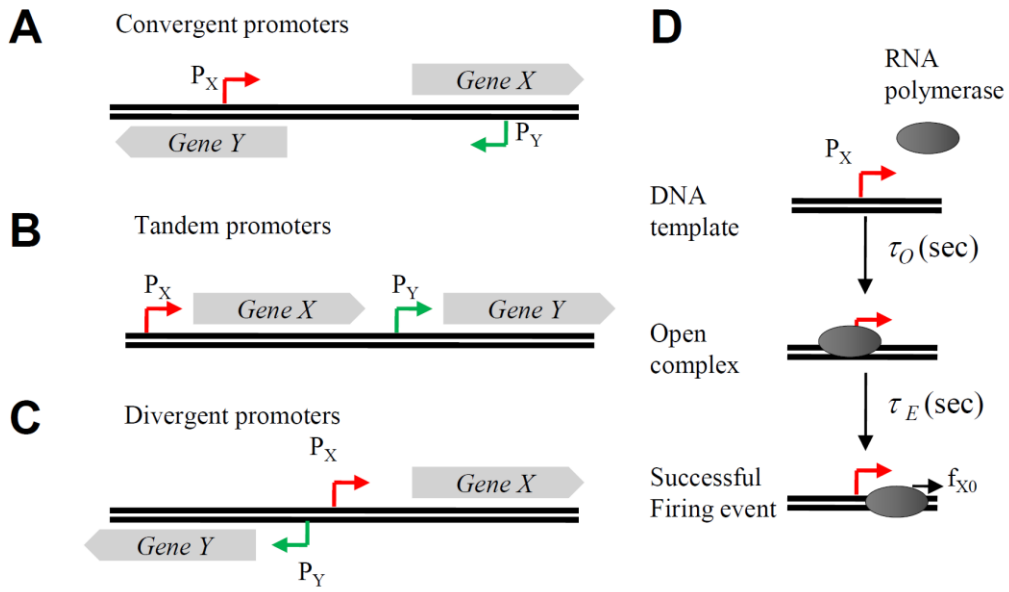


Figure 2.11: Different promoter arrangements that lead to transcriptional interference. A pair of promoters P_X and P_Y driving expression of genes *X* and *Y* are shown in different arrangements, namely, **(A)** Convergent promoters, **(B)** Tandem promoters and **(C)** Divergent promoters. **(D)** Two steps involved in RNAP: promoter interaction. (i) RNAP binds at a promoter at certain time interval τ_0 and (ii) the closed promoter complex transitions into an open promoter complex at time intervals τ_E once the initiation complex is formed.

2.7 Transcriptional Interference

Transcriptional interference (TI) is defined as the suppressive influence of one transcriptional process on an adjacent transcriptional process occurring *in cis* due to RNA polymerase traffic along the DNA (Shearwin *et al.*, 2005). TI occurs when a strong promoter reduces the expression of an adjacent weak promoter. TI has been reported in a

number of studies on both prokaryotic (Callen *et al.*, 2004a; Palmer *et al.*, 2009; Ward and Murray, 1979) and eukaryotic systems (Greger *et al.*, 2000; Gullerova and Proudfoot, 2008; Hongay *et al.*, 2006; Prescott and Proudfoot, 2002). TI has also been reported in causing HIV virus latency(Coiras *et al.*, 2009; Lenasi *et al.*, 2008), influencing transcription of cancer related genes (Sara *et al.*, 2009) and influence transcription depending on site of chromosomal integration (Eszterhas *et al.*, 2002).

Promoters are arranged in three possible orientations shown in Figure 2.11 A-C. Transcriptional interference occurs when binding or movement of RNA polymerase at or from one promoter is influenced by the activity of the adjacent promoter. The mechanism by which transcriptional interference operates is broadly classified into four categories, which are described in the following. The act of RNAP firing from a promoter comprises of two main steps: (i) formation of an open complex, and (ii) formation of elongation complex. Both these steps influence the final RNAP firing time interval from a promoter and determine its strength (Fig 2.11 D).

2.7.1 RNA polymerase collision

In this mechanism, transcriptional interference occurs due to collision between elongating RNAP polymerases originating from convergent promoters (Figure 2.12 A). For successful transcription, RNA polymerases need to traverse a length of overlapping DNA shared by both genes. Transcription from such a locus results in two possibilities, either elongating RNA polymerases continue elongating in absence of a converging RNA polymerase from the opposing promoter, or converging RNA polymerases encounter

each other, collide and fall-off the DNA. Atomic Force Microscopy imaging of *in vitro transcription* convergent promoter templates have shown that converging RNA polymerases collide (Crampton *et al.*, 2006). While a significant fraction of collided RNAPs fall off the DNA, however, it was observed that often one of collided RNAPs continued to bind to the DNA after a brief pause (Crampton *et al.*, 2006). Such collision results in premature termination of transcriptional progress of one or both complexes. Studies on the *gal7* and gal10 genes of *S.cerevisiae* showed that arranging genes in convergent orientation decreased transcription from this region due to increased RNAP collision.

2.7.2 Promoter Occlusion

Promoter occlusion occurs when transcription from one promoter prevents RNAP from binding at the other promoter due to steric hindrance (Figure 2.12 B). This is due to limitation in the time available for an RNAP to bind to a promoter due to transient passage of an RNAP from the opposing promoter. This can occur for both tandem as well as convergent promoters. In the case of tandem promoters, when RNAPs from an upstream promoter fail to terminate, these can interfere with a downstream promoter by displacing RNAPs or other DNA-binding proteins that may lie in their path. Sitting duck is more prevalent the two interfering promoters are located to close to each other and that one of the promoters is much more aggressive than the other.

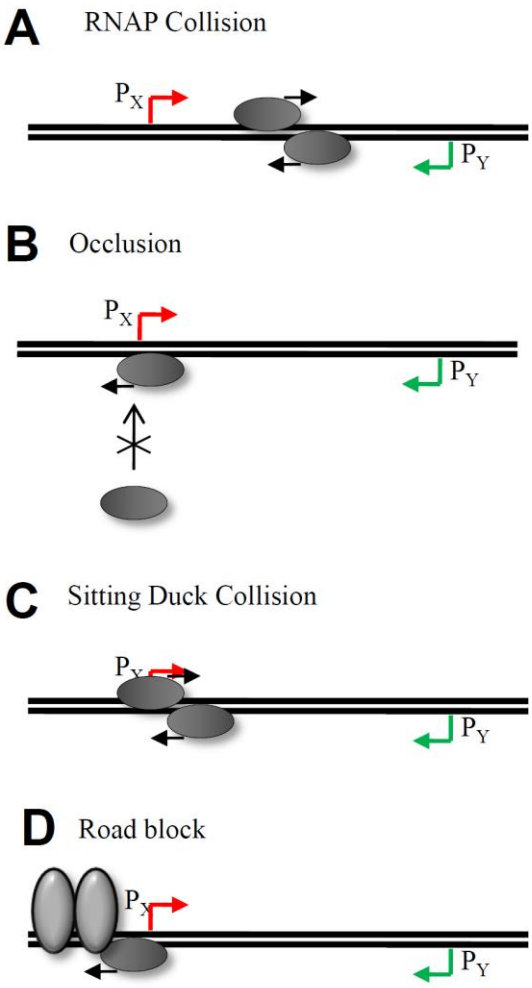


Figure 2.12: Mechanisms of transcriptional interference. A pair of convergent promoters, namely P_X (sense strand, indicated by red) and P_Y (antisense strand, indicated by green) are considered to exert transcriptional interference via four mechanisms. **(A)** RNAP collision, **(B)** Promoter occlusion, **(C)** Sitting Duck Collision and **(D)** Roadblock.

2.7.3 Sitting Duck Collision

An RNAP, stationed at a weak promoter, due to slow transition between an closed promoter complex to an open promoter complex, serves as a "sitting-duck" available for collision with an elongating RNAP from an aggressive promoter (either convergent or tandem) as shown in Figure 2.12 C. Again, this mechanism is more pronounced in cases

when promoters are separated by shorter distances, such as the pR-pL convergent promoter system separated by 62 bp in coliphage 186 (Callen *et al.*, 2004) or the *scbA-scbR* system separated by 53 bp in *Streptomyces coelicolor* (Chatterjee *et al.*, 2011). Often RNAPs pause during movement along DNA, in some case, such paused complexes serve as sitting-ducks for collision as shown in PR-PRE system in bacteriophage λ (Palmer *et al.*, 2009).

2.7.4 Road blocking

Road blocking effect occurs when DNA bound proteins serve as a roadblock or obstacle to the movement of RNAP along the DNA (Figure 2.12 D). This could be caused by tightly bound open promoter complexes or large DNA bound transcription factors.

2.8 Antisense RNA interaction

Bacteria possess numerous regulatory RNAs which regulate gene-expression by either regulating transcription, or via post-transcriptional mechanisms such as affecting translation or by interacting with proteins or base pairing with target RNAs etc (Waters and Storz, 2009). In the past decade, there has been burst of novel small RNAs identified across different bacterial species, a majority have been characterized to interact via base-pairing with other RNAs (Thomason and Storz, 2010), many of these participate in regulatory networks (reviewed in (Beisel and Storz, 2010)). Bacterial regulatory RNAs

which act via base-pairing can be broadly into two classes: *cis*-encoded and *trans*-antisense RNAs (Brantl, 2007). The *trans*-encoded RNAs are encoded from distinct chromosomal locations and share limited base-complementarity. On other hand, *cis*-encoded antisense RNAs are encoded on the opposite strand of the DNA and have the potential to form extensive base-pairing interactions with corresponding sense RNAs.

Antisense RNAs can vary in size from tens to thousands of nucleotides, however, a large number of antisense RNAs characterized fall in the shorter size range (~50 to 300 nt). During transcription, as RNAP reads through the DNA, the nascent RNA starts folding into complex secondary structures due to base-pairing of local sequence (Figure 2.13 A). A large number of antisense RNAs have been found to be non-coding, highly-structured and interact with target sense RNAs via single-stranded region along the RNA such as via stem-loop structures or bulges (Brantl, 2007). Typically, interaction between sense and antisense RNA have been shown to occur between 5-8 nt G-C rich loops in a 5'-YUNR motif which is proposed to form a U-turn structure causing a sharp bend in the RNA phosphate backbone, providing a platform for rapid interaction with complementary RNA (Franch *et al.*, 1999).

The resulting hybrid RNA complex between sense and antisense RNA is subjected to degradation or rendered untranslatable (Brantl, 2002a; Brantl, 2007). Small antisense-RNAs have been reported to play a regulatory role in other bacterial systems, including 108 nt RNAI RNA controlling copy number of plasmid ColEI (Tomizawa and

Itoh, 1981), 69 nt Sar RNA of bacteriophage 22 repressing Ant protein (Liao *et al.*, 1987), 77 nt OOP RNA of bacteriophage λ

repressing CII protein (Krinke and Wulff, 1987), 90 nucleotide CopA RNA controlling copy number of plasmid R1 (Stougaard *et al.*, 1981) and 70 nt Sok RNA repressing post-segregation killing in plasmid R1 (Thisted and Gerdes, 1992). Three of the major mechanisms by which antisense RNA causes post-transcriptional inhibition are described in the following.

2.8.1 Transcriptional attenuation

Antisense interaction between a nascent sense RNA and antisense RNA results in transcriptional termination of the sense RNA. In systems when transcriptional attenuation operates, the sense RNA can adopt two mutually exclusive conformations depending on presence or absence of antisense RNA. Interaction of the nascent sense RNA with antisense RNA induces a terminator stem-loop, which consequently results in premature transcriptional termination. In absence of antisense RNA, the sense RNA folds in to what is known as the 'anti-terminator' configuration which allows transcriptional read-through and ultimately protein synthesis. Such a regulatory role of antisense RNA has been observed in a number of systems, including, interaction between Anti-Q and nascent Q RNA of pCF10 plasmid system in *Enterococcus faecalis* (Johnson *et al.*, 2010), *rep* mRNA if staphylococcal plasmid pT181(Brantl, 2002) and RNAIII/RNAII of plasmid pIP501 (Brantl, 2002) (Figure 2.13 B).

2.8.2 Translational inhibition

The most common mechanism of action used by antisense RNA is by direct blocking of ribosomal binding site of the sense RNA to prevent its translation into proteins. Such a mechanism has been reported for *traJ*-FinP duplex of plasmid F and R1 (Figure 2.13 C), where, FinP antisense RNA block the RBS of *traJ* which encodes for activator of the conjugative operons of plasmid F and R1 (Brantl, 2002). A similar role of antisense RNA in down regulating synthesis of proteins from *cis*-encoded genes has been reported in a number of systems, including, *Sok* mRNA of plasmid R1 in *E.coli* (Thisted and Gerdes, 1992), antisense RNA *alr1690-furA* regulating expression of transcriptional repressor FurA in cyanobacterium *Anabena* sp. PCC 7120 (Hernandez *et al.*, 2006) and other systems reviewed in (Thomason and Storz, 2010).

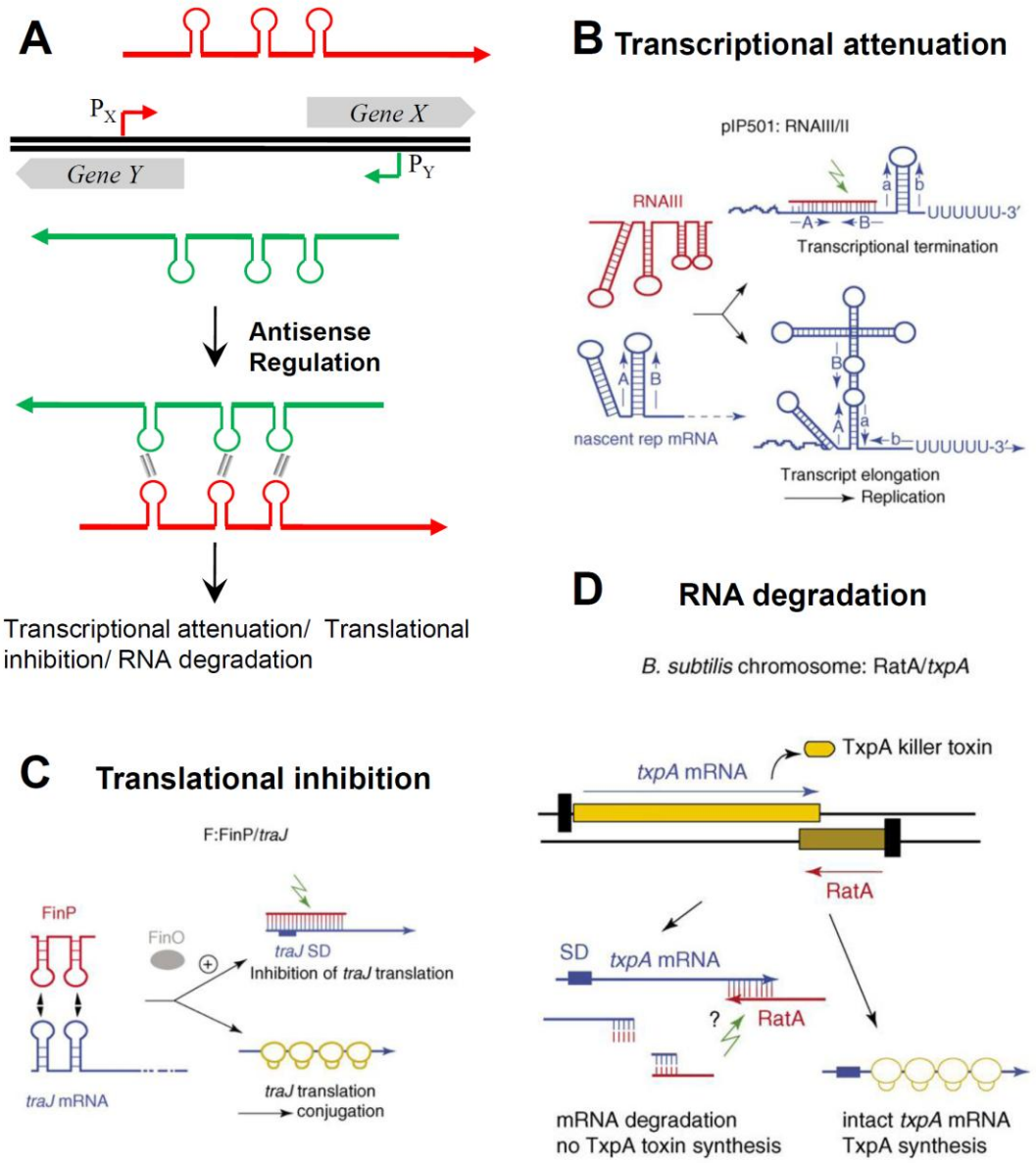


Figure 2.13: RNA interaction between *cis*-encoded antisense RNAs. (A) A general mechanism of interaction between sense and antisense RNA, the duplex is a target for transcriptional attenuation, translational inhibition or RNA degradation. Examples of sense and antisense RNA interaction in different biological systems that demonstrate **(B)** transcriptional attenuation, **(C)** translational inhibition and **(D)** RNA degradation. Figures (B-D) are courtesy (Brantl, 2007).

2.8.2 RNA degradation

The third major mechanism, by which antisense RNA regulates sense RNA expression, is via influencing the stability of sense RNA. RNAs typically have short half-lives and are continuously removed from the cells by the action of RNA nucleases (RNases). Interaction with antisense RNA, can cause structural changes which render the sense RNA more vulnerable to RNases. For example, the RatA/TxpA antitoxin/toxin system from *Bacillus subtilis* employs an antisense RNA interaction mediated mRNA degradation, which is considered to be a result form RNaseIII-cleavage of RatA/*txpA* duplex (Silvaggi *et al.*, 2005) as shown in Figure 2.13 D.

CHAPTER 3

TWEAKING BIOLOGICAL SWITCHES THROUGH A BETTER UNDERSTANDING OF BISTABILITY BEHAVIOR

3.1 Summary

This chapter reviews the concept of bistability in the context of biological regulatory networks. Many biological events are binary. The switch between mutually exclusive OFF to ON state in response to a stimulus is frequently mediated by a control circuit with a positive and/or a negative feedback. Such a system typically exhibits hysteresis with its switching ON and OFF stimulus levels dependent on the current state of the system. The system can be shown to be bistable both experimentally and mathematically. Work to synthesize such switches by combining natural or engineered components has begun to illustrate the potential of such control circuits in many areas of applications.

3.2 Introduction

Many biological processes incur "states" which are binary, cells or the organisms exist only in one of the mutually exclusive states. For example, a microbial cell may be either a vegetative cell or a spore at a given time, but not both. A cell can be either in or not in an apoptotic state; a pluripotent stem cell may be at an undifferentiated state capable of differentiating to multiple lineages and replicate, or at a state committed to differentiating to a particular lineage. At biochemical levels, glycolysis and gluconeogenesis, with their reaction fluxes in opposite directions, are mutually exclusive metabolic states. Numerous such biological switches occupy critical decision-making points in the process of life. Failure in their control action is often catastrophic. In the course of evolution, these switches have become highly robust. However, from a technological perspective, it might be desirable to alter the control characteristics of some switches such as in the case of directed differentiation of stem cells. In other applications such as gene therapy one may wish to exert control on the on-set of gene delivery through a synthetic switch. A better understanding of the mechanism of the switches will lead to their wider applications in biotechnology.

Mathematical models were also used to demonstrate the role of bistability in sonic hedgehog (Shh) signaling pathway, which plays an important role during development (Lai *et al.*, 2004). The positive feedback loop and negative loop embedded in the system endows it with two distinct states in response to Shh. Stochastic simulations have demonstrated that the feedback loop may dampen the fluctuations of the positive

autoregulator (Gli) and render the system robust. In another study, the behavior of Oct4-Sox2-Nanog, the three transcription factors important in determining the fate of embryonic stem cells, was examined theoretically with a number of putative network structures (Chickarmane *et al.*, 2006). The stability behavior of various network structures may aid in the experimental investigation of stem cell fate determination.

3.3 Network dynamics and bistability behavior

A key characteristic of a genuine switch is that, in response to a concentration change of the stimulus over a threshold value, the system turns from one state to the other; the system does not reside in a state that is in between the two states. The machinery that makes up a biological switch usually entails a network of genes and their transcriptional, translational and biochemical products. The participation of those constituents in a series of protein-protein, protein-DNA and other molecular interactions as well as biochemical reactions results in a switch behavior (Figure 3.1). The response of individual reactions or molecular interactions with respect to the concentration of the participating component is continuous and is usually graded (Figure 3.1 B). However, the combination of those individual ramping responses gives rise to the sharp, binary (ON-OFF) type of switch behavior.

In principle, an ON-OFF type of response can also be accomplished by a single chemical reaction such as one with a highly cooperative behavior (one whose kinetics is described by a Hills function with a large exponent). However, such a system is extremely sensitive to fluctuations in stimulus concentration. Upon switching to an ON

state after crossing a threshold stimulus concentration, the system will quickly return to an OFF state if the stimulus concentration fluctuates below the threshold level. Such a behavior is unacceptable for a robust switch. Studies in the past few years have led to the emergence of some general features for such ON-OFF switches in various biological systems. The network of these switches generally consists of a positive feedback and a negative feedback (Figure 3.1 A) that can be shown mathematically to exhibit a bistable behavior. These general features and bistability are described below.

A robust switch behaves as though it possesses a memory of the signal. Once at an ON state, even if the stimulus concentration is reduced to below the threshold level, it continues at the ON state. In some special cases, the system stays ON even if the stimulus is completely removed. Xiong et. al. (Xiong and Ferrell Jr., 2003) demonstrated such irreversibility in cell fate decision making during the maturation of *Xenopus* oocytes. The maturation of *Xenopus* oocytes is mediated by the phosphorylation and dephosphorylation of p42 mitogen activated protein kinase (MAPK) and cell-division protein kinase Cdc2. These kinases are activated during oocyte maturation and are responsive to steroid stimulus. It was elegantly demonstrated that after being turned on upon exposure to a high stimulus concentration, p42 MAPK and Cdc2 activities are maintained even if stimulus was removed.

In a bistable system, the threshold concentration of the signal required for the system to be switched from OFF to ON state is different from that for transitioning in the reverse direction. If one performs an experiment by adding stimulus to the system in

small increments, and assesses the state of the system, one would obtain a response curve (C-D-A in Figure 3.1 C). Reversing the direction of the experiment, one generates another response curve. The downward threshold concentration differs from that of the upward. Such a behavior is referred to as hysteresis. In such a system, the switch of the state is not sensitive to fluctuations in stimulus concentration. Many experimental systems have shown hysteresis ((Bagowski and Ferrell, 2001; Bagowski *et al.*, 2003; Laslo *et al.*, 2006; Ozbudak *et al.*, 2004; Pomerening *et al.*, 2003b; Sha *et al.*, 2003)).

At a population level, the switch between mutually exclusive states is often graded in time as well as in stimulus concentration. However, at a single cell level the response is still ON-OFF. Stochastic behavior among different individuals in a population is inevitable, resulting in different individual cells responding at somewhat different concentrations and time. When observed at a population level, the response is thus graded. Such contrast between population level and single cell level has been illustrated experimentally in a number of systems, including *Xenopus levis* oocytes (Bagowski and Ferrell, 2001; Bagowski *et al.*, 2003), and oxidative stress induced apoptosis mediated by ERK pathway and p53 pathway (Nair *et al.*, 2004).

The control circuit or network structure of a number of systems with switch behavior has been formulated in mathematical forms and their behavior analyzed. Typically, the system of interest is confined to a small segment of cellular reactions, and the system is considered as isolated, thus neglecting the interactions with other cellular regulatory circuitry and material flows. Usually, a stability analysis is performed by

allowing the system to reach a steady state and then the system is allowed to bifurcate. After the introduction of perturbation, if the system returns to the original steady state, then the point is considered stable. If it drifts out of the original steady state point, the system is not stable. If the system is in a stable region introducing a small change of the stimulus concentration to a new level this will lead the system to a new steady state point corresponding to the new stimulus concentration. The steady state analysis of the switch system often leads to a behavior depicted in Figure 3.1 D. In the two segments, the lower curve containing points C and D and the upper curve containing points B and A, are stable steady states, whereas the region of the curve between B and D is unstable. In the region enclosed by ABCD thus there are three steady states for each stimulus concentration, two are stable and one is unstable. If the system is originally at an ON state (Point A, Figure 3.1 D), and the concentration is increased gradually, the system moves gradually along the steady state line. Upon reaching Point B, further increase in stimulus concentration will shift the system to the new point C in the other stable region which is at an OFF state. Further increase in stimulus will continue in the stable OFF region. Reversing the trend of stimulus concentration change will cause the system to return to point C.

3.4 Experimental observation of bistability

Bistability has also been observed during cellular differentiation in bacteria. A notable example is antibiotic production in *Streptomyces coelicolor* (Mehra *et al.*, 2008). Many members of streptomycetes species use a pheromone, γ-butyrolactone to

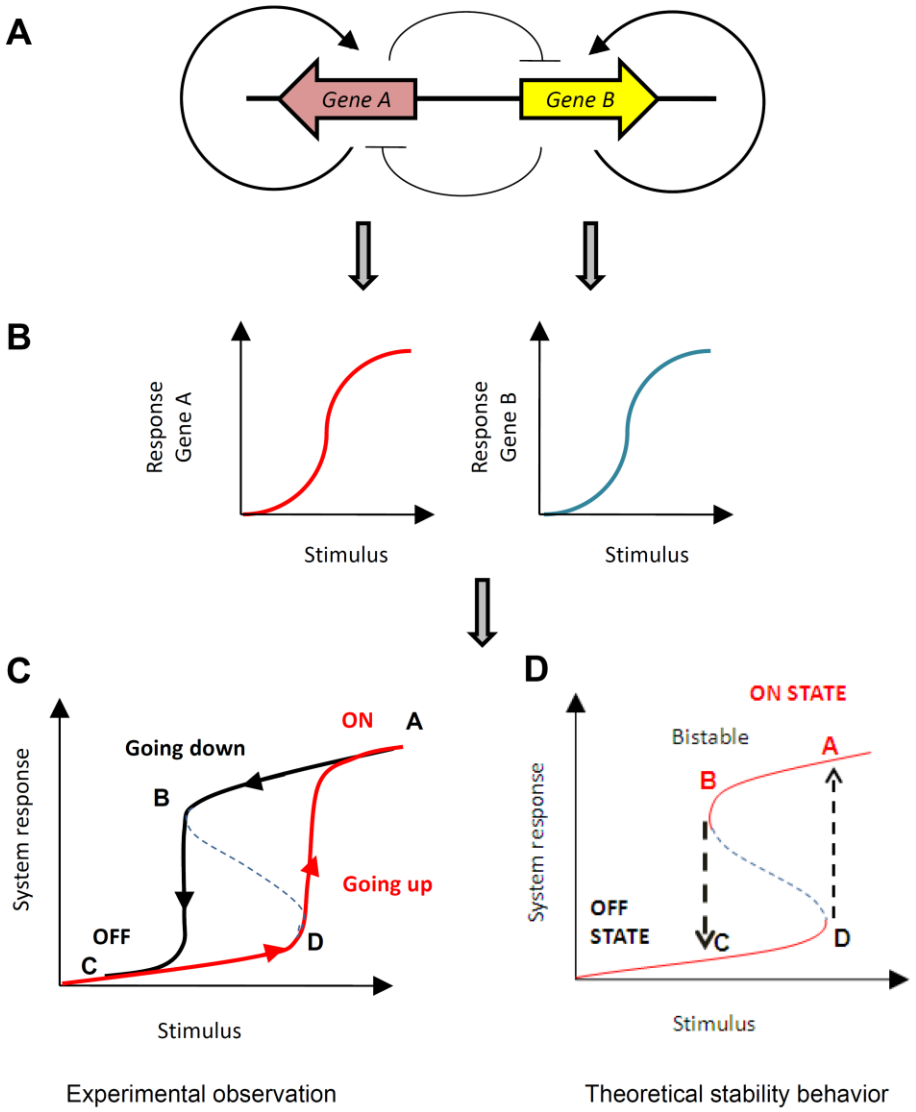


Figure 3.1: Defining bistability in context of a two-gene regulatory network. (A) A two gene network with positive and negative feedback loops. **(B)** Graded response at individual gene level. **(C)** System demonstrating hysteresis, the threshold concentration for switching the system from OFF to ON state (point D) is higher than that required to switch the system from reverse direction (point B). **(D)** Bistable response at the system level. Region enclosed within ABCD has three steady states for each concentration of stimulus. Two are stable and experimentally observable. The one unstable one (dashed line) can be theoretically demonstrated.

synchronize the population for antibiotic production and sporulation. The regulatory cascade in *coelicolor* includes two genes and their products, an autorepressor protein (ScbR) and a signal amplifier protein (ScbA) which synthesizes the signaling molecule.

The repression by ScbR is neutralized by the formation of ScbA/ScbR complex and the binding of signaling molecule Scb1 to ScbR. Thus Scb1 suppresses the negative feedback of ScbR by removing the active form of ScbR (indicated by dash line in Figure 3.2 A). ScbA/ScbR complex is postulated to be positive feedback for ScbA. The delicate balance of ScbA and ScbR affects their own expression (Figure 3.2 A). As a result, when a few cells in a population start releasing Scb1, the derepression of ScbA causes rapid synthesis of signaling molecules. The accumulation of Scb1 triggers the production of antibiotics. It also triggers others in the population to synthesize Scbl, thus synchronizing the entire population for antibiotic production. In *S. coelicolor* the corresponding genes of ScbA and ScbR, namely *scbA* and *scbR*, were shown to transiently increase in their expression and then subside to a low level (Mehra *et al.*, 2008). A mathematical model for the butyrolactone system in *Streptomyces* species that describes the dynamics of the system was shown to exhibit a bistable switch behavior, wherein a threshold concentration of signaling molecule switches the system from a stable OFF (no antibiotic) state to a stable ON state (antibiotic production). It was articulated that antibiotics are not only growth inhibitory to other microbes in their surroundings, but also to the producing cells. Producers express antibiotic resistance machinery before the on-

set of antibiotic production. Thus, a switch type of control that also synchronizes the population is critical for the producing cells.

Growing cells oscillate between interphase and mitosis of the cell cycle. The switch from interphase to mitosis is controlled by an oscillatory regulatory network, consisting of positive and negative feedback loops. In the early embryo of *Xenopus*, cdc2-cyclin B functions as an oscillator (Figure 3.2 B). Cdc2 activates the Anaphase promoting complex (APC), which then degrades cyclin B. The cyclin B is involved in activating Cdc2 by forming the Cdc2/cyclin B complex. The active Cdc2/cyclin B activates the cdc25 which further activates Cdc2. It is also involved in inactivating the inhibitors wee1 and Myt1. The system thus entails negative feedbacks, positive feedbacks, and a mechanism to remove the negative feedback like in the ScbA-ScbR system of *S. coelicolor*.

Pomerening et. al. (Pomerening *et al.*, 2003) demonstrated that these positive feedback loops impart the system bistability. They demonstrated that the activation of Cdc2 with respect to cyclin B exhibits hysteresis. This allows the cell to settle in either the M-phase or interphase but not in any intermediate state. Using extracts from unfertilized *Xenopus eggs* in M-phase that can be forced into interphase by destruction of cyclin B, and then manipulating the concentrations of cyclin B has shown that two distinct levels of cdc2 activity exit in the 45-60 nM range of cyclin concentration. The level of cdc2 activity is dependent on whether the system was going up from interphase

to M-phase or vicecersa. This provided experimental evidence that the system exhibits hysteresis, a characteristic of bistability.

Bistable switch is also frequently observed in determining cell fate during development. Before lineage commitment, multipotential hematopoietic progenitors express low levels of multi-lineage genes (Laslo *et al.*, 2006). The fate of the multi-potent cell is determined mainly by the actions of two primary transcription factors, PU.1 and C/EBPα, (Figure 3.2 C). The expression of PU.1 causes an increase in transcription of Egr/Nab which activates the expression of genes that lead to macrophage development. On the other hand, expression of C/EBPα activates the transcription of Gfi which activates the transcription of genes that lead to a neutrophil lineage. Both Egr/Nab and Gfi repress each other. Expression of PU.1 prior to C/EBPα, leads to macrophage cell fate. However, induction of C/EBPα was before PU.1, results in neutrophil cell fate. The hysteresis in the system prevents the differentiated cell to switch to the other state. Through a mathematical model for the gene regulatory network that successfully simulated the experimental observation the system was shown to be bistable and is robust to perturbations. The bistable state can potentially be exploited to obtain a single (monostable) differentiated state by shifting the balance between Egr/Nab-2 and Gfi-1 thereby guiding the cell to commit to a particular lineage.

3.5 Mathematical analysis of bistability

The existence of bistable steady states can give rise to switch behavior. However, it is not clear that all biological switches are explained by bistability. For many systems,

detailed experimental studies on their bistable behavior is difficult. Many have taken a mathematical approach to demonstrate the possible existence of bistability. It is often argued that a biologically viable system should not be sensitive to parameter value change. Since biologically cells reside in an environment where virtually all biological parameters fluctuate around a mean value, any biological switch must function in spite of those noises. There is no consensus on how best to evaluate the robustness of the response to multiple parameters. Testing each parameter over a range of value independently does not predict the behavior when a number of parameters change simultaneously. The large combination of different parameters over their biologically feasible range of values can be a daunting task. Nevertheless, demonstration of the robustness of a system's bistable behavior can give credence to its being a biological switch.

The decision of a cell to undergo apoptosis has been thought to behave like a switch (Wee and Aguda, 2006). A mathematical model was used to demonstrate the role of pro-apoptotic Bax and anti-apoptotic Bcl-2 in determining the apoptotic cell fate (Bagci *et al.*, 2006). A highly cooperative formation of apoptosome was shown to give a sudden increase in the concentration of the apoptotic signal (caspase-3) in response to a slight increase in the concentration of the input stimuli. The ratio of Bax to Bcl-2 is critical in deciding the cell fate. Therefore, controlling the ratio of Bax: Bcl-2 by manipulating their degradation or synthesis rate could direct the cell to either survival or apoptosis. With the help of reduced models that included cooperative interactions and a

positive feedback loop, the cooperative binding and the formation of the apoptosome was shown to play a key role in imparting bistability to the model. To assess the robustness of the system, different combinations of values of rate constants were tested using the reduced model. Twenty-six of the 81 combinations tested showed bistable behavior. However, with the feedback loop alone and without cooperative effect no bistable behavior was observed.

3.6 Synthetic gene switches

The switching characteristic of gene expression in response to stimuli has aroused interest in synthetic biology, an emerging field that uses well-characterized gene modules to *de novo* construct gene networks with desirable system characteristics (Guet *et al.*, 2002). Genetic devices such as an epigenetic toggle switch and oscillators in *E.coli* have been reported (Elowitz and Leibler, 2000; Gardner *et al.*, 2000; Hasty *et al.*, 2002). Synthetic biologists are gradually turning their attention to building eukaryotic (Becskei *et al.*, 2001) and in particular mammalian synthetic gene networks (Greber and Fussenegger, 2007; Kramer *et al.*, 2004; Kramer and Fussenegger, 2005). By constructing a gene network containing two antibiotic inducible transcription control systems (streptomycin and macrolide responsive, respectively), that are mutually repressive to each other, a synthetic system was shown to toggle between HIGH and LOW states (Kramer et. al., 2004). It was also shown that once a state was achieved, it could be maintained even in the absence of antibiotics for several generations. A synthetic mammalian switch (Kramer and Fussenegger, 2005) was constructed with two

opposing components. One provides a positive feedback with a self-inducing tetracycline responsive transactivator (VP16). The other component consists of a Erythromycin (EM) responsive transrepressor (E-KRAB), which represses the positive feedback in the presence of EM. These two opposing modules impart the system with its hysteretic behavior. Ajo-Franklin et. al. (Ajo-Franklin *et al.*, 2007) constructed a high fidelity memory device in yeast based on a transcriptionally controlled positive feedback. There have been many theoretical approaches to synthetic biology (Sotiropoulos and Kaznessis, 2007). Novel networks have been engineered by recombining existing genetic parts. In fact, by systematically investigating different components of the network one can rationally design novel gene networks.

3.7 Outlook of manipulating stability for metabolic engineering

With switch type of regulatory structures playing pivotal roles in development, cell fate determination and alteration of physiological state, it is natural that biotechnologists target those switches for metabolic engineering, especially for *in vivo* applications in higher organisms or even in humans. To date examples of such manipulation of natural control circuits are still few, however, synthetic switches that are capable of switching ON and OFF under different induction conditions have begun to emerge. Possible applications of such synthetic switches are numerous. Examples include, controllable delivery of medicines in gene therapy, cell fate control in cell therapy and event-triggered protein expression in plants for enhanced protection against adverse environmental conditions.

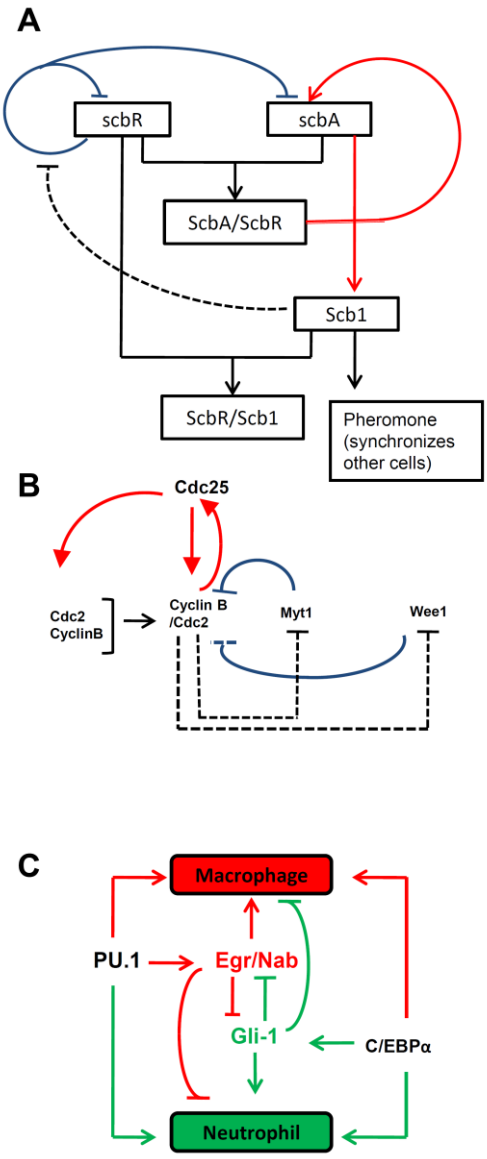


Figure 3.2: Examples of biological genetic-networks where bistability is observed. (A) Schematic diagram of the ScbA/ScbR system regulating antibiotic production in *Streptomyces coelicolor*. Arrows represent positive regulation and blunt arrows represent repression. Red and blue colors denote positive and negative feedback respectively. The brown color indicates negation of repression control and the dash line indicates negative feedback directed toward a negative feedback. ScbR acts both as an autorepressor as well as a repressor of ScbA. The signaling molecule Scb1 (stimulus) derepresses ScbR repression of itself and ScbA by binding to ScbR. Depression of ScbA causes Scb1sysnthesis, causing further synthesis of ScbA and Scb1. The response of ScbR to stimulus Scb1 has been shown to have a bistable behavior. **(B)** Cell cycle hysteresis observed in *Xenopus* oocyte cell maturation. The genetic network involves a number of positive feedback loops which impart the system hysteresis. The positive feedback loops include activation by dephospohorylation of Cdc2/cyclin by Cdc25. Cdc2 is involved in the activation of Cdc25, which in turn activates Cdc2 and inactivates Myt1. Myt1 is an inhibitor of Cdc2. (Adapted from Pomerening et al, 2003). **(C)** Differentiation of hematopoietic stem cells from a multi-lineage gene expression state to two different lineages, macrophage and neutrophil. The primary cell fate determinants are PU.1 and CEBP/α which are expressed constitutively. They are involved in induction of both macrophage specific and neutrophil specific genes. Secondary cell fate determinants Egr/Nab and Gfi-1 are responsible for inducing macrophage and neutrophil specific genes respectively. (Adapted from Laslo et al, 2006).

As our understanding of the mechanism of switch behavior increases, the means of manipulating the bistability dynamics will also expand. It is conceivable that an antagonist or protagonist can be used to modulate the interactions of key control elements and change the bistability behavior in cell development without resorting to genetic manipulation. So far the synthetic switches are largely based on genetic elements of bacterial origin. Employing control elements of higher organisms will be essential for many applications. More significantly, one needs to have a better assessment of the robustness of the natural and synthetic switches. How sensitive are natural switches to fluctuations in the levels of the components of the circuit? How sensitive is the switch behavior to alteration in the strength of molecular interactions? In most cases, the sensitivity or robustness of the system cannot be easily assessed experimentally.

The biological switches examined so far all have a relatively simple structure. For bacterium, they often consist of only a pair of genes, their gene expression and biochemical product. Yet the interactions of those few elements can give rise to robust and sophisticated switch behavior. A systems approach certainly offers the best way of understanding how the complexity of switches arises from the simplicity of gene structure and the best way of helping us develop means of tweaking the stability behavior. In marching toward the goal of manipulating biological switches, one sees the synergism between systems and synthetic biology. On one hand while systems biology emphasizes a top-down approach to elucidate regulatory components of particular gene network, synthetic biology on the other hand focuses on a bottom-up approach to

construct artificial genetic circuits for producing desirable system characteristics. We hope this synergism will unlock many opportunities and create another fertile ground for innovation.

CHAPTER 4

CONVERGENT TRANSCRIPTION CONFERS A BISTABLE SWITCH IN ENTEROCOCCUS FAECALIS CONJUGATION

4.1 Summary

Convergent gene pairs with head-to-head configurations are wide spread in both eukaryotic and prokaryotic genomes and are speculated to be involved in gene regulation. Here we present a novel mechanism of gene-regulation due to convergent transcription from the antagonistic *prgX/prgQ* operon in *Enterococcus faecalis* controlling conjugative transfer of the antibiotic resistance plasmid pCF10 from donor cells to recipient cells. Using mathematical modeling and experimentation, we demonstrate that convergent transcription in the *prgX/prgQ* operon endows the system with the properties of a robust genetic switch through pre-mature termination of elongating transcripts due to collisions between RNA polymerases (RNAP) transcribing from opposite directions, and antisense regulation between complementary counter transcripts. Evidence is provided for the presence of truncated RNAs resulting from convergent transcription from both the promoters that are capable of sense-antisense interactions. A mathematical model predicts that both RNAP collision and antisense regulation are essential for a robust

bistable switch behavior in the control of conjugation initiation by *prgX/prgQ* operons. Moreover, given that convergent transcription is conserved across species, the mechanism of coupling RNAP collision and antisense interaction is likely to have a significant regulatory role in gene expression.

4.2 Introduction

Convergent transcription, from two opposing promoters of partially overlapping genes on opposite strands of DNA gives rise to transcripts with potential sense-antisense interactions in the overlapping region. Such convergent transcription is widespread in eukaryotes including the mouse (Katayama *et al.*, 2005), human (Yelin *et al.*, 2003), *Drosophila melanogaster* (Misra *et al.*, 2002), and *Arabidopsis thaliana* (Yamada *et al.*, 2003) genomes and a number of prokaryotes such as *L. monocytogenes* (Toledo-Arana *et al.*, 2009), *M. pneumoniae* (Guell *et al.*, 2009), *E.coli* (Dornenburg *et al.*, 2010). We postulate that convergent, overlapping gene organization provides two-layered regulation by a combination of transcriptional interference and antisense regulation through RNA: RNA interactions between complementary transcripts (Brantl, 2007).

Transcription from promoters of convergent overlapping genes gives rise to a finite probability that opposing elongating RNA Polymerases (RNAP) collide head-on, thus exerting a suppressive effect on transcription. Such suppressive influence of one transcriptional activity on a second transcriptional activity occurring *in cis* is referred to as transcriptional interference (Shearwin *et al.*, 2005), as has been reported in prokaryotic (Callen *et al.*, 2004b; Palmer *et al.*, 2009) and eukaryotic systems (Greger *et al.*, 2000a;

Napoli *et al.*, 2009). Furthermore, the two transcripts, having complementary sequence in the overlapping region, may exert antisense regulation through RNA: RNA interactions (Brantl, 2007; Waters and Storz, 2009). Given the frequent occurrence of sense-antisense transcripts in both prokaryotic and eukaryotic genomes, convergent transcription may play a significant role in gene regulation by both mechanisms. However, there have been few attempts to document the relative effect of each mechanism in control of transcription of overlapping, convergent genes.

The present study investigates the role of transcriptional interference arising due to RNAP collision and antisense regulation due to expression of complementary transcripts from operons *prgQ* and *prgX*, in plasmid pCF10 that regulate the conjugative transfer of antibiotic resistance in plasmid pCF10 between *Enterococcus faecalis* donor and recipient cells (Dunny, 2007). Promoter P_Q drives expression of the *prgQ* operon, which encodes pheromone inhibitor peptide, iCF10 (AITLIFI) as well as the pCF10 conjugation machinery (Nakayama *et al.*, 1994). Approximately 223 bp downstream of *prgQ* start site, the convergent promoter P_X drives expression of the *prgX* operon which encodes repressor PrgX regulating P_Q transcription by binding to the operators XBS 1 and 2 of promoter P_Q (Figure 4.1 A-C). In addition to the full length *prgX* transcript (denoted as X), transcription from P_X also produces a 104 nt non-coding RNA, Anti-Q, by processing or termination within the 5' end of the *prgX* transcript (Bae *et al.*, 2002). The interaction of Anti-Q with nascent *prgQ* transcripts affects folding of *prgQ* RNA and

results in a short 380 nt terminated Q_S transcript (at IRS1, Figure 4.1 A) which encodes iCF10, but is incapable of inducing conjugation (Johnson *et al.*, 2010).

P_Q is normally in a repressed state while P_X is constitutive (Figure 4.1 B). The switch of P_Q between repressed state (conjugation OFF, Figure 4.1 B) and derepressed state (conjugation ON, Figure 4.1 C) is controlled by competing interactions of PrgX with inhibitor iCF10 and an inducer, cCF10, a chromosomally- coded heptapeptide (LVTLVFV) that is produced predominantly by recipient cells (Nakayama *et al.*, 1994) since in donor cells, the pCF10 encoded PrgY protein sequesters endogenous cCF10 (Buttaro *et al.*, 2000; Chandler *et al.*, 2005). In the absence of recipient cells and thus low ratios of cCF10 to iCF10, iCF10 bound PrgX tetramers cause a DNA loop between XBS 1 and 2 sites, stabilizing a repressing complex at P_Q (Figure 4.1 B) (Dunny *et al.*, 1978; Shi *et al.*, 2005). In the derepressed state, the cCF10 imported from the environment via membrane protein PrgZ (Leonard *et al.*, 1996) displaces iCF10, resulting in the disruption of PrgX-iCF10 tetramers and the DNA loop (Figure 4.1 C). This leads to increased levels of *prgQ* transcription, which serves to titrate the Anti-Q RNA, and results in expression of a longer 530 nt transcript called Q_L (Figure 4.1 A), capable of inducing conjugation by a post-transcriptional mechanism (Bensing *et al.*, 1997).

In this report we employed a mathematical model to demonstrate that convergent transcription of *prgX/prgQ* system plays a regulatory role in endowing the system a bistable switch-like behavior, a characteristic of many robust biological switches, including regulation of competence in *Bacillus subtilis* (Maamar and Dubnau, 2005),

lysogeny and lytic state in bacteriophage λ (Tian and Burrage, 2004) and during cell fate determination in *Xenopus levis* oocytes (Pomerening *et al.*, 2003). We provide experimental evidence that RNAP collision gives rise to shorter truncated transcripts which exert antisense effect on transcripts of *prgX* and *prgQ*. Both transcriptional interference as a result of RNAP collision and antisense regulation are necessary for the *prgX/prgQ* genetic switch to function in a robust manner.

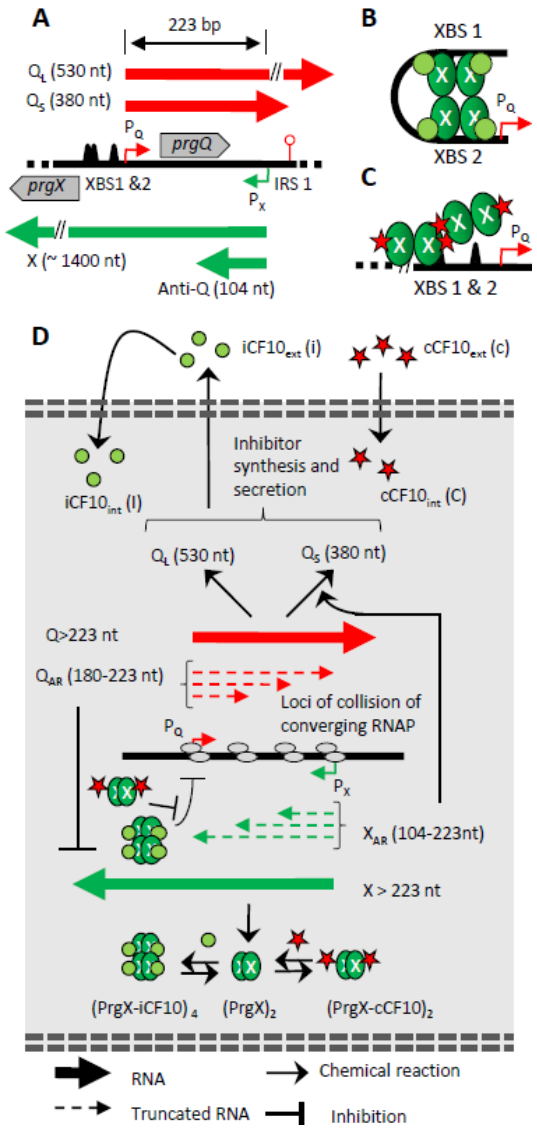


Figure 4.1: The *prgQ-prgX* genetic locus controls conjugation of pCF10 plasmid in *Enterococcus faecalis*. (A) Convergent promoters P_Q (red) and P_X (green) drive expression of the antagonistic *prgQ* and *prgX* operons respectively with a 223 bp overlap. Transcripts (indicated by bold arrows) from the inducible P_Q promoter include 380 nt Q_S RNA terminating at an Inverted Repeat Sequence (IRS1) and 530 nt Q_L RNA which results in expression downstream conjugation-related genes. Transcripts from constitutive antisense P_X promoter include non-coding Anti-Q and full-length X RNA coding for PrgX protein. **(B)** Promoter P_Q in repressed state and **(C)** derepressed state. **(D)** Model for the *prgQ-prgX* genetic switch. Successful transcription gives rise to full length Q and X RNA (>223 nt). Failed transcription, upon collision of converging RNAP, gives rise to various sizes of truncated RNA (<223 nt) indicated by dashed arrows. Some truncated RNA (Q_{AR} and X_{AR}) have the potential to exert antisense interaction with full length counter transcripts**. (Not to scale)**

4.3 Results

4.3.1 Modeling prgQ-prgX genetic network: coupled effect of RNAP collision and Antisense Regulation

A model for conjugation control by the *prgX/prgQ* operons is shown in Figure 4.1 D. The proposed model is based in part on previously described mechanisms (Dunny, 2007b; Johnson *et al.*, 2010). We included additional regulatory effects contributed by transcriptional interference caused by RNAP collision and antisense effects caused by truncated RNA produced as a result of transcription interference. The underlying hypothesis is that RNAP collision causes failure in transcription, thus suppressing the transcription rate from both promoters. A second element is the generation of truncated transcripts which are released upon RNAP collision. We consider the hybridization of these short RNAs to their complementary counter-transcripts.

A two-part mathematical model was formulated to describe the mechanism and to quantitatively evaluate the behavior of transcription from P_Q and P_X in response to inducer cCF10 (Shearwin *et al.*, 2005). The first part is a discrete model which enumerates the frequency of RNAP collision and the generation of truncated RNA. The second part is an ordinary differential equation (ODE) based mathematical model on the mass-balance of the components of the pCF10 system (Figure 4.1 D).

The discrete model simulated RNAP moving along the overlapping DNA between P_Q and P_X (details in Chapter 5 and Figure 5.1). Upon RNAP binding to the promoter, a delay of 10s is allowed for formation of elongation complex (Tang *et al.*, 2009) before

RNAP moves at a velocity of elongation at 50 bp/s, which has been reported as the average velocity of RNAP in *E.coli* in presence of pausing (Neuman *et al.*, 2003). The maximal RNAP binding rate at a derepressed P_Q was set at 0.1/s based on RNAP initiation rates estimated for P_{bla} system in *E.coli* (Liang *et al.*, 1999) and PR-PRE system in bacteriophage λ (Palmer *et al.*, 2009). Using experimentally determined relative strengths of promoters P_Q and P_X under repressed and derepressed states in Figure 4.2, the RNAP binding rates at P_Q and P_X under these two states were calculated. RNAP collision is considered to occur when converging RNAPs from P_Q and P_X are separated by 60 bp (RNAP footprint). This takes into consideration occlusion of P_Q due to passage of RNAP originating from P_X and vice versa.

Our experimental data suggested that RNAP collision did not result in complete abolition of transcription based on a 10% level of X transcript observed upon derepression, we thus assume that 90% of RNAP collisions result in transcriptional termination (Crampton *et al.*, 2006), releasing short transcripts terminated at the point of collision (described in Chapter 5 and Figure 4.7) . Movement of RNAP is not hindered by binding of PrgX to DNA (Figure 4.10 A). Successful passage of RNAP across the overlapping region (223 nt) without collision generates the full length Q or X RNA. RNAP collision gives rise to a population of truncated RNA with varying length of less than 223 nt from P_Q and P_X (Figure 4.1 D). Previous work showed that stem loops at 46-51 nt and 80-83 nt of P_X transcript respectively interact with stem loops at 173-178 nt and 156-161 nt of P_Q transcript (Figure 4.3, for sequence see Figure 4.4 A) (Shokeen *et al.*,

2010). Based on this data and *in silico* RNA structural analysis using Sfold (Ding and Lawrence, 2003), we postulate that truncated P_X RNAs (including Anti-Q RNA) between 104-223 nt in length and truncated Q RNA between 180 nt and 223 nt are capable of RNA: RNA interaction with their counter transcripts Q and X respectively due to presence of stem loops (Figure 4.8). This sub-population of truncated RNAs capable of antisense effects are collectively denoted as Q_{AR} (180-223 nt) and X_{AR} (104-223 nt) RNA from P_Q and P_X promoters respectively (Figure 4.1 D).

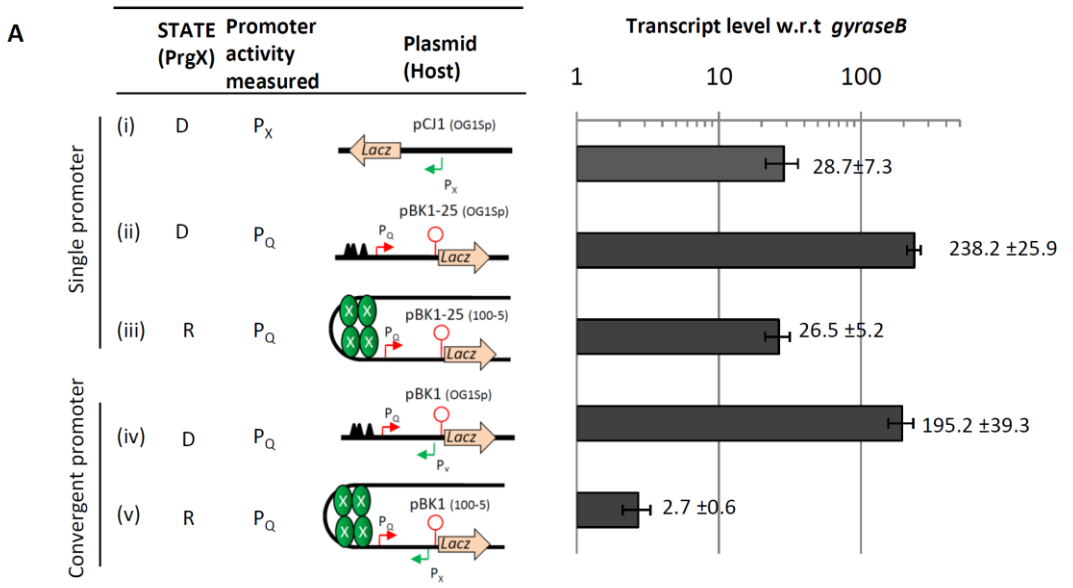
From the discrete model simulation the rate of generation of Q, X, Q_{AR} and X_{AR} species were calculated (Figure 4.5 A-B) for simulation using the ODE model (Chapter 5, Figure 5.1). A set of 8 ODE's describe balance of RNA species Q_S, Q_L, X, X_{AR} and Q_{AR}, intracellular iCF10 and cCF10 and extracellular iCF10 (Equations 5.25-5.32) considering the rate of production, degradation and dilution due to volume expansion caused by growth as well as interactions among these components (see Chapter 5). For RNA species, interaction between sense transcripts (Q and Q_{AR}) and antisense transcripts (X and X_{AR}) following second order kinetics (Shokeen *et al.*, 2010) was also considered (see Chapter 5). Hybrid RNA duplex between X RNA and Q_{AR} is considered to be unavailable for translation (Figure 4.1 D) (Brantl, 2007). The formation of RNA duplex between X_{AR} and nascent Q RNA gives rise to formation of Q_S RNA (Johnson *et al.*, 2010; Shokeen *et al.*, 2010). Nascent Q transcripts that do not terminate proceed to elongate and form Q_L. Translation of Q_S and Q_L leads to production of the secreted inhibitor peptide iCF10 (Dunny, 2007).

Transport of iCF10 and cCF10 across the membrane protein PrgZ is considered to follow first order kinetics (Leonard *et al.*, 1996). The model assumes that due to cCF10-sequestering activity of PrgY protein, the effect of endogenous cCF10 in donor cells is negligible (Chandler *et al.*, 2005).

The ODE model was solved to obtain steady state solution to a fixed concentration of extracellular cCF10 (Figure 4.5 C-D). The parameter values were obtained from the literature or experimental data (Table 5.1 and Figure 4.8). To test the sensitivity of the steady state behavior to the parameter values, the parameters in the model were lumped into six dimensionless parameters (Table 5.2) to reduce the model to a set of 5 non-dimensional ODE's (Equations 5.48-5.52). The values of three of these dimensionless numbers were determined from experimental data (Chapter 5), whereas the remaining three were obtained from literature and subjected to sensitivity analysis (Table 5.4). In the dimensionless model, the term corresponding to RNA interaction between truncated RNA's X_{AR} and Q_{AR} was eliminated as it did not affect the steady state behavior of the system. Simultaneous parameter space search for the remaining 3 dimensionless parameters demonstrated that steady state behavior shown in Figure 4.5 C-D existed over at least two orders of magnitude (Figure 5.2, Table 5.4). The details of model derivation are described in Chapter 5.

4.3.2. P_Q activity in presence and absence of convergent transcription

The strength of P_Q in the presence (repressed state) and absence (derepressed state) of repressing PrgX complexes was quantified by measuring steady state expression levels of the *lacZ* transcript from plasmids containing either single P_Q promoter or both convergent promoters with *lacZ* reporter fused downstream of IRS1 (Figure 4.2). Two kinds of host strains were used: (1) OG1Sp, wildtype host strain, and (2) 100-5, a derivative of OG1Sp which provides PrgX *in trans* from the chromosome at levels similar to wildtype pCF10 plasmid (Figure 4.4 B). By combining the plasmids and host strains a set of five conditions were created as shown in Figure 2. The intrinsic promoter strength of P_Q (derepressed state) in absence of convergent transcription is 8.3 fold higher than the constitutive P_X (Figure 2, ii vs. i). The repression of P_Q by supplying PrgX *in trans* caused a 9 fold reduction in expression in the absence of convergent transcription (iii vs ii). In the repressed state, P_Q has almost identical strength to P_X (i vs iii). In the derepressed state P_Q becomes the stronger promoter (i vs ii). While convergent transcription from P_X caused only a modest (~15%) reduction of transcripts from P_Q in the derepressed state (stronger P_Q than P_X, ii vs iv), it reduced P_Q transcript levels by about 90 fold in the repressed state (similar strength of P_Q and P_X, ii vs v). These results show that convergent transcription greatly enhances the effect of repression on P_Q.



B

Ratio of *LacZ* expression	Derepressed state	Repressed state
$\frac{P_Q\ (single)}{P_X\ (single)}$	8.3 (ii/i)	0.92 (iii/i)
$\frac{P_Q\ (single)}{P_Q\ (convergent)}$	1.22 (ii/iv)	9.8 (iii/v)

Effect of repressor on expression from promoter P_Q	9 (ii/iii)
Combined effect of repression and interference from convergent transcription	88.2 (ii/v)

Figure 4.2: Interference exerted on P_Q expression due to convergent transcription. (A) Promoter activity of P_Q and P_X was estimated by measuring *lacZ* transcript level by qPCR normalized to *gyraseB* for the constructs indicated. The activity of derepressed P_Q (denoted by D) was assessed in strain OG1Sp which lacks *prgX*, while that of repressed P_Q (denoted by R) was measured in the isogenic strain 100-5, which expresses PrgX from a constitutive *prgX* allele integrated into the chromosome. Five conditions were used for the analysis: **(i)** pCJ1 (OG1Sp): Single promoter construct for measuring expression from native P_X, **(ii)** pBK1-25 (OG1Sp): Single promoter construct for measuring expression from derepressed P_Q, **(iii)** pBK1-25 (100-5): Single promoter construct for measuring expression from repressed P_Q, **(iv)** pBK1 (OG1Sp): Convergent promoter construct for measuring expression from derepressed P_Q in the presence of P_X, and **(v)** pBK1 (100-5): Convergent promoter construct measuring expression from repressed P_Q in the presence of P_X. The *lacZ* mRNA expression levels shown are an average of three independent qPCR experiments (error bars are standard deviations from mean values). **(B)** Summary of data in (A)

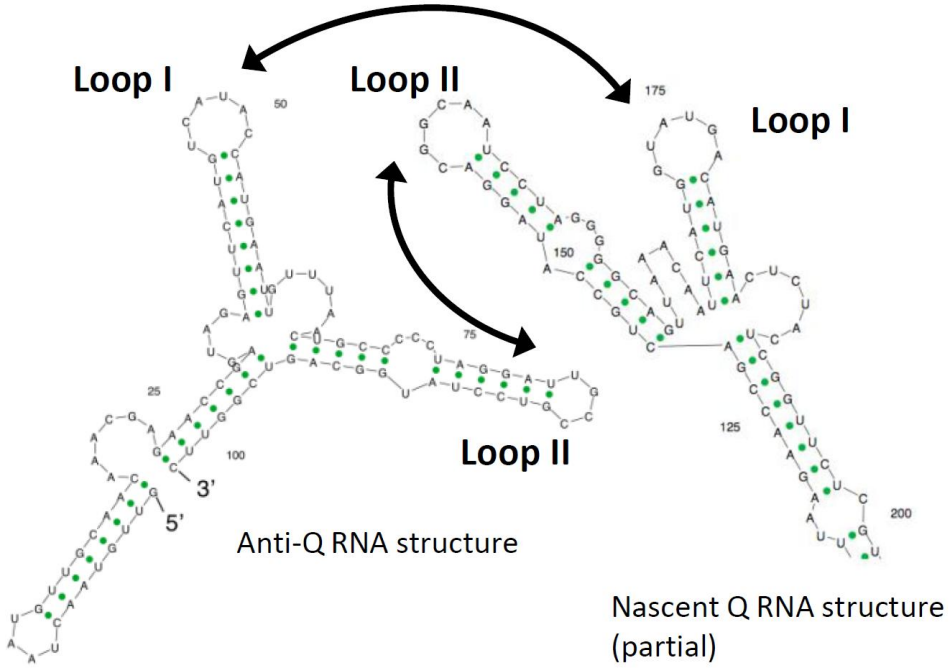


Figure 4.3: Regions of interaction between *prgQ* and Anti-Q transcripts. Loops I and loop II of Anti-Q (Loop I: 46-51 nt and Loop II: 80-83) and Q_S (Loop I: 173-178 and Loop II: 156-161) transcripts (partial structure shown) have been demonstrated to interact (Shokeen *et al.*, 2010).

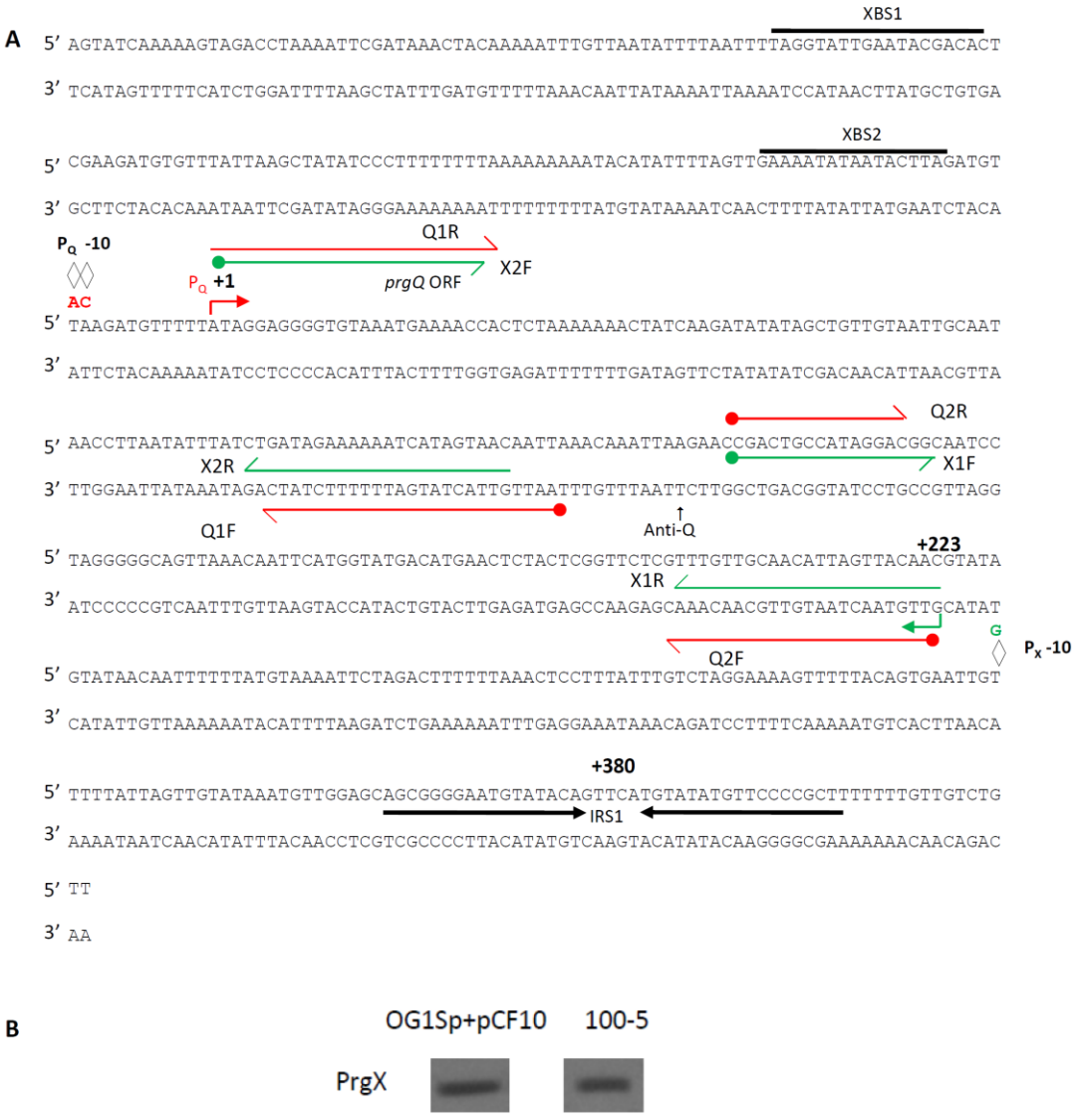


Figure 4.4: Sequence map for *prgQ-prgX* locus and Western blot for PrgX expression. **(A)** The primers (forward/reverse) used for designing probes Q1 (Q1F/Q1R), Q2 (Q2F/Q2R), X1 (X1F/X1R) and X2 (X2F/X2R) have been indicated. Point mutations in -10 regions of P_Q and P_X for inactivating respective promoters are indicated by ◊. The PrgX binding sites are shown as XBS 1 and XBS 2. The positions +1, +223 and +380 w.r.t. P_Q promoter are indicated. **(B)** Western Blot showing similar levels of PrgX from plasmid pCF10 in OG1Sp and 100-5 strain.

4.3.3 Model simulation and bistable behavior

The discrete model was used to predict the relative abundance of transcripts under repressed and derepressed conditions as a result of RNAP collision (Figure 4.5 A-B, Table 5.1). Under repressed condition, higher numbers of transcription events are predicted to give rise to full length X RNA, compared to the depressed condition, whereas the opposite is predicted for full length Q RNA. On the other hand, a broader distribution of Q_{AR} and X_{AR} RNA is predicted in the repressed state compared to the derepressed state. In the derepressed state, the 8.3-fold stronger P_Q not only increases Q transcription but also knocks off most RNAP initiated from P_X at loci proximal to the P_X promoter. As a result, RNAP collision results in only a moderate decrease in Q transcription, but a rather large decrease in X transcription.

Using the ODE model we next examined the effect of RNAP collision and antisense regulation on the steady state level of Q_L in response to the signaling molecule cCF10. The level of Q_L RNA is an indicator of the state of conjugation, since it encodes genes of conjugation machinery. Steady state solution of the model shown in Figure 4.5 C demonstrates a characteristic bistable response of Q_L to cCF10 concentration. Multiple steady states reside in the S-shaped section of the curve. Two are stable steady states corresponding to ON (upper) and OFF (lower) states respectively. The unstable steady state in the middle is not observed experimentally. For the system at an OFF state, as cCF10 concentration increases it moves along the lower stable steady state (Q_L <1.4 nM) until cCF10 reaches 3.7 ng/mL, then the system undergoes a sharp transition to an ON

state (conjugationally competent) corresponding to high level of Q_L (>3.8 nM). Conversely if the system is initially at an ON state, it remains at an ON state till cCF10 concentration decreases below 3.2 ng/mL where it rapidly changes to an OFF state. The system is thus marked by a well separated ON (cCF10 > 3.7 ng/mL) and OFF state (cCF10< 3.2 ng/mL).

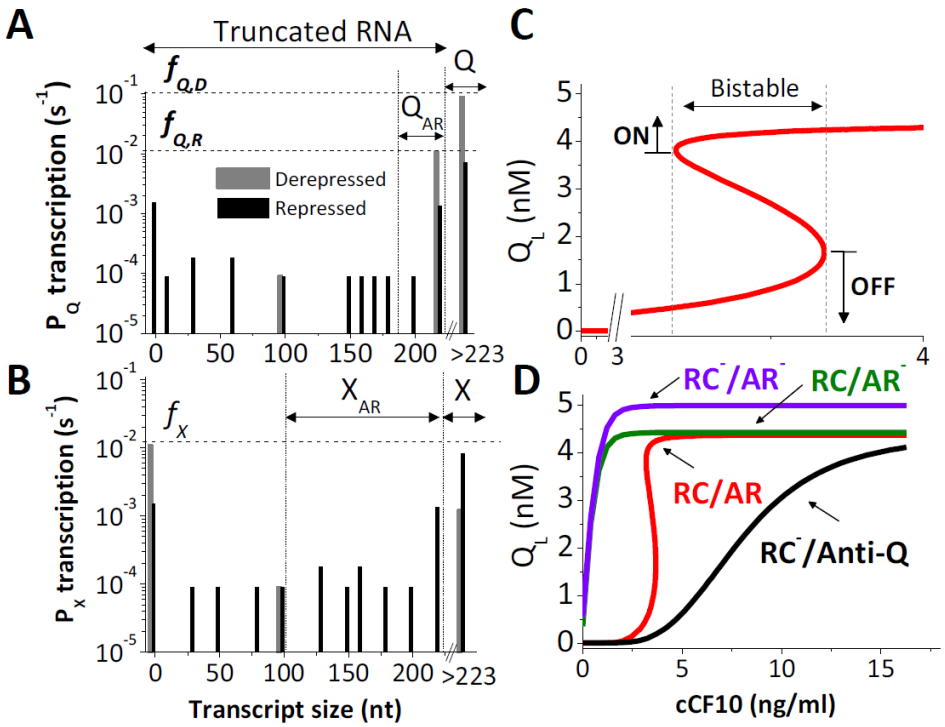


Figure 4.5: Mathematical modeling predicts RNAP collision and antisense regulation are required for bistability in the *prgQ-prgX* switch. (A-B) Frequency of various sizes of truncated and full length RNA as a result of RNAP collision in the repressed (RNAP firing rates $f_{Q,R}$=0.011 s^{-1} from P_Q, and f_X= 0.012 s^{-1} from P_X) and derepressed ($f_{Q,D}$=0.1 s^{-1}, f_X= 0.012 s^{-1}) state from P_Q **(A)** and P_X **(B)**. **(C)** Steady state level of Q_L at different cCF10 concentrations shows characteristic bistable switch behavior. **(D)** Effect of de-coupling RNAP collision (RC) and antisense regulation (AR) on the steady state response of Q_L to cCF10.

We evaluated the relative contribution of RNAP collision and antisense regulation to bistable behavior by eliminating either or both RNAP collision (RC) and antisense regulation (AR) from the model (Figure 4.5 D). AR was eliminated from the model by setting the rate constant for RNA interaction to zero, whereas, RC was eliminated by setting transcription rate of Q and X RNA to RNAP firing rates from P_Q and P_X respectively and X_{AR} and Q_{AR} concentration to zero (Table 5.3-5.4). Bistability is observed only when both RC and AR are present (RC/AR). When AR is absent while RC is present (RC/AR^-) or both RC and AR (RC^-/AR^-) are eliminated, no bistability is seen. We also examined the case where RC is only partially removed by reducing the RNAP falling off rate upon RNAP collision. As the falling-off rate decreases to 50% the steady state behavior transitions from a bistable to a ramp-like response (Figure 4.7 A). The results thus hinted that the 104 nt Anti-Q RNA, produced as a result of X RNA processing, alone is insufficient for the bistable behavior. This is indeed the case for RC^- /Anti-Q, the bistable response to cCF10 is lost and a ramp-like transition is observed (Figure 4.5 D).

4.3.4 Inverse expression and response to pheromone

We next measured the levels of transcripts from P_X and P_Q promoters in strain OG1Sp carrying wildtype (WT) plasmid pCF10. Under repressed conditions Q_L transcript was much lower than X transcript even though the strength of P_Q and P_X are similar. Upon induction with high levels of cCF10 (50 ng/mL), the Q_L transcript level increased more than 200 fold, while the X transcript decreased approximately 9 fold

(Figure 4.6 A). The opposite trend in their response to derepression is consistent with model predictions (Figure 4.5 C and Figure 4.7 B). The level of Q_L RNA observed was lower than that seen in the *lacZ* reporter system (Figure 4.2). This could be attributed to a high degradation rate for Q_L. This notion is consistent with model simulation, where a 10 fold decrease in Q_L degradation rate abolishes bistable behavior (Figure 4.6 D).

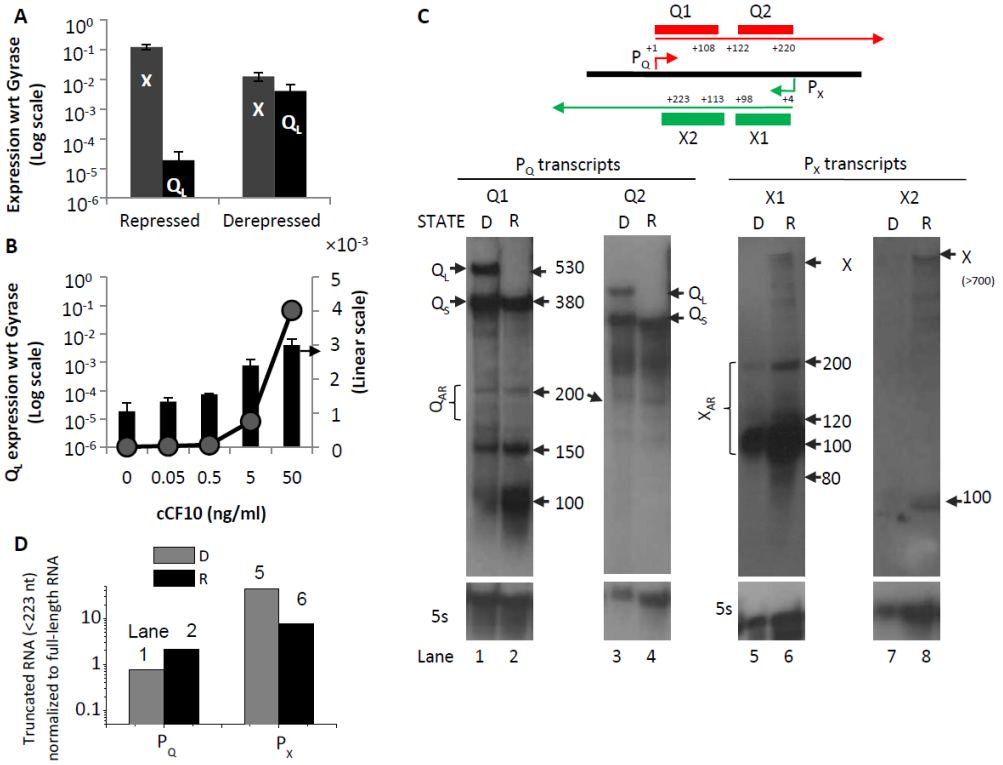


Figure 4.6: Inverse relationship of Q_L and X expression and presence of truncated RNAs in the *prgQ-prgX* locus. (A) Q_L and X RNA levels measured by qPCR using X-specific and Q_L-specific primers from OG1Sp (pCF10) under repressed (0 ng/mL cCF10) or derepressed (50 ng/mL cCF10) conditions. **(B)** Levels of Q_L RNA in response to different concentrations of cCF10. Data shown in (A) and (B) are averages of three independent experiments (error bars are standard deviations from mean values). **(C)** Top: Map for the single stranded RNA probes Q1, Q2, X1 and X2 used for Northern blotting experiments. Bottom: Northern blots to detect transcripts from P_Q and P_X in cells carrying WT pCF10, in either derepressed (D) or repressed (R) state. Probes/lanes: Q1/1-2, Q2/3-4, X1/5-6 and X2/7-8. **(D)** Quantitative estimation of intensity of bands for truncated P_Q and P_X transcripts in lanes 1-2 and 5-6 normalized to intensity of Q_S+Q_L and X respectively using ImageJ software (NIH).

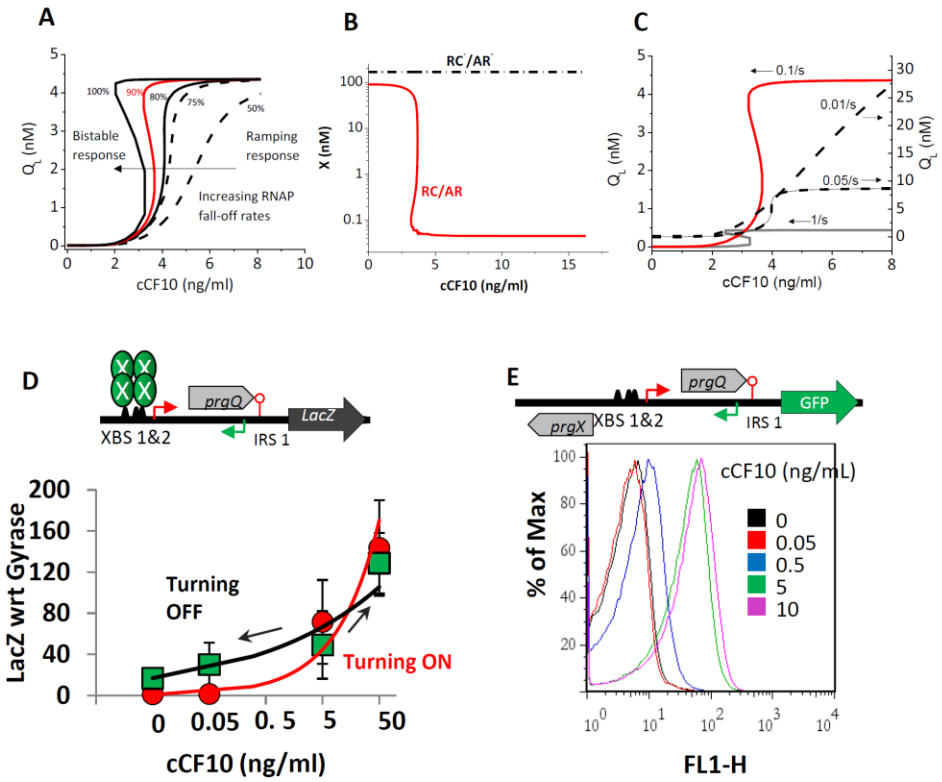


Figure 4.7: Bistable behavior of *prgQ-prgX* gene switch. (A) Steady state response of Q_L to cCF10 for different RNAP fall-off rates during RNAP collision. **(B)** Inverse steady state expression of X (w.r.t Q_L) in response to cCF10 for RC/AR case (constant steady state expression of X for RC^-/AR^- case). **(C)** Bistable response of Q_L to cCF10 is lost at lower degradation rates of Q_L. Steady state Q_L level for higher degradations rates corresponding to 0.1/s and 1/s are shown on the primary y-axis (left), for lower degradations rates corresponding to 0.05/s and 0.01/s are shown on the secondary y-axis (right). **(D)** Convergent transcription in strain 100-5 with pBK1 plasmid (contains convergent P_Q-P_X region of pCF10 with *lacZ* reporter fused downstream of IRS1 and PrgX being provided constitutively from the chromosome) induced with different concentrations of cCF10 (0, 0.05, 5 and 50 ng/ml) for 60 min to reach steady state corresponding to the turning ON behavior (red line). Cells induced at 50 ng/ml of cCF10 (60 min) were washed and exposed to different concentrations of cCF10 (0, 0.05, 5 and 50 ng/ml) for 60 min to reach steady state corresponding to the turning OFF behavior (black line). Levels of *lacZ* expression were normalized to *Gyrase B*. Data shown is an average of three independent experiments (error bars are standard deviations from mean values). **(E)** Flow cytometric data showing GFP expression as an indicator of ON state. GFP was fused downstream of P_Q promoter in a convergent promoter construct (pBK2Δ*lacZ::gfpmut3b*) (Kozlowicz, 2005) which excludes other factors present on WT pCF10 plasmid. Cells containing this plasmid were induced with different concentrations of cCF10 (0, 0.05, 0.5, 5 and 10 ng/ml) for 60 min after which steady state was reached. Cells transition to ON state at concentrations between 0.5 and 5 ng/ml of cCF10.

A dose/response curve of Q_L transcript to cCF10 showed that Q_L expression remains low below 0.5 ng/mL of cCF10; followed by a 8.3 fold increase at 5 ng/mL of cCF10 (Figure 4.6 B and Figure 4.7 D) characteristic of a switch from OFF to ON state as predicted by the model (Figure 4.5 C). Flow cytometric data using GFP protein expression as an indicator of ON state, confirms that the system transitions from an OFF to ON state between 0.5 and 5 ng/mL (Figure 4.7 E).

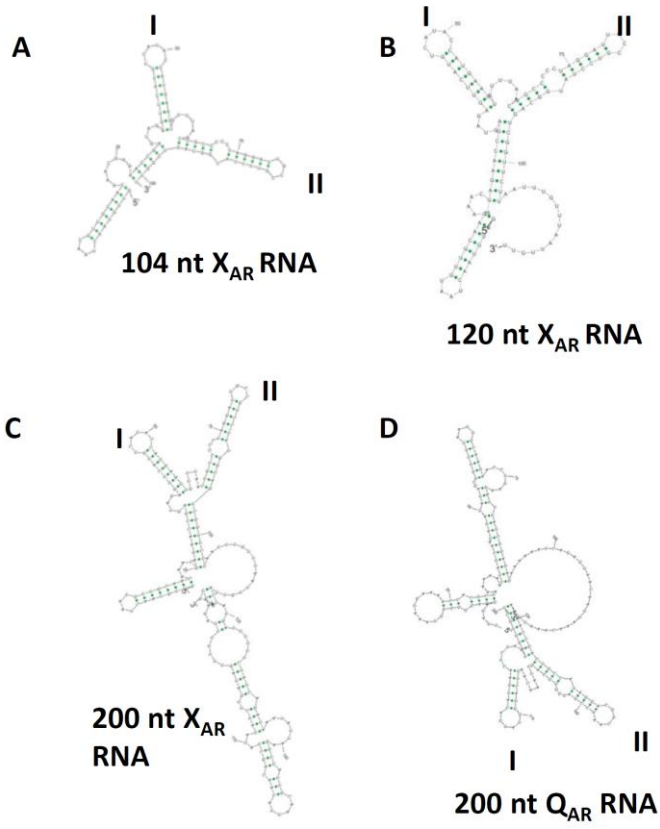


Figure 4.8: Regions of interaction between *prgQ* and Anti-Q transcripts. (A-C) *In silico* structures of X_{AR} (A-C) and Q_{AR} (D) truncated RNA species corresponding to sizes between 180-200 nt and 104-223 nt as observed in wildtype pCF10 (Figure 4.6 C) generated using Sfold software are shown (Ding and Lawrence, 2003). Regions involved in RNA interaction have been indicated by stem loops I and II.

4.3.5 Truncated RNA from P_Q and P_X

To verify the existence of the truncated transcripts (X_{AR} and Q_{AR}) in pCF10 carrying donor (OG1Sp) cells, we used Northern blots with sets of probes complementary to the 5' (Q1 and X1) and 3' segments (Q2 and X2) within the overlapping region of the respective RNAs from each promoter (Figure 4.6 C, top panel). Shorter (truncated) P_Q (Figure 4.6 C, lane 1-4) and P_X (lane 5-8) transcripts within the overlapping region were indeed observed under both repressed and derepressed conditions. The Q1 probe detected truncated transcripts of ~100, 150 and 200 nt sizes, with higher relative abundance in repressed cells (lane 2 vs. lane 1, Figure 4.10 C). Similarly, X1 probe detected truncated RNAs in the size range between 80 to 200 nt (lane 5-6).

The prominent band of approximately 100 nt corresponds to the 104 nt Anti-Q RNA reported previously (Bae *et al.*, 2002). The level of truncated transcripts detected by probes Q1 and X1 (lanes 1-2 and 5-6) proximal to 5' end are higher than those by distal probes Q2 and X2 (lanes 3-4 and 7-8), consistent with the higher collision frequency near the 5' end. The presence of a short 100 nt band detected by X2 probe (lane 8) indicates an additional mechanism involved in processing of longer P_X transcripts. However, this truncated transcript is unlikely to have a regulatory role in the switch behavior, due to absence of stem loops I and II required for antisense interaction with P_Q transcripts. The high abundance of truncated RNA in the repressed state (lane 2 vs 1 and 6 vs 5, Figure 4.10 C) is consistent with the prediction of the RNAP collision model (Figure 4.5 A-B). The size of shorter RNAs observed fall within the range of Q_{AR} and X_{AR} considered in the model and contain the structure to participate in antisense interactions (Figure 4.8). However, it is possible that these short

RNAs arise from degradation of full-length transcripts. In this case, the ratio of truncated RNA to full length RNA would be similar under derepressed and repressed conditions. Conversely if these are the products of RNAP collision, we expect this ratio to increase for P_X transcripts in the derepressed state as X RNA decreases, while the ratio corresponding to P_Q transcripts should decrease because of dramatic increase in Q (sum of Q_S+Q_L) transcripts upon induction. Quantification of northerns in Figure 4.6 C (lane 1-2 and 5-6) indeed show that the ratio of truncated to full-length RNA increases for P_X transcripts and decreases for P_Q transcripts in the derepressed state compared to the repressed state (Figure 4.6 D).

4.3.6 Decoupling RNAP collision and Antisense Regulation

The relative contribution of RNAP collision and antisense regulation to switch response was evaluated by decoupling these effects using single or convergent promoter constructs in host strains OG1Sp and 100-5 for derepressed and repressed conditions respectively (Figure 4.9 A). Three cases were analyzed: convergent transcription (RC/AR), no RNAP collision in absence of antisense RNA (RC^-/AR^-) and presence of antisense RNA supplied *in trans* (RC^-/AR) using plasmid constructs shown in Figure 4.9 A.

Transcription from P_X in the presence of convergent transcription (RC/AR) showed abundant truncated RNA from a repressed P_Q compared to derepressed P_Q (Figure 4.9 B, lane 2 vs 1). The longer (>700 nt), likely X transcript, was more abundant in the repressed condition, consistent with our model prediction (Figure 4.5 B, Figure 4.7 B). The distribution of transcripts was somewhat different from Figure 4C because of

different combinations of host cells and plasmid constructs. The absence of convergent transcription (RC^-/AR^-) resulted in increased expression of both Anti-Q RNA (~100 nt band) and longer X transcripts (Figure 4.9 B, lane 3), as compared to the case with convergent transcription (lane 1), consistent with the negative effects of RNAP collision and antisense regulation on P_X expression. A 200 nt band observed in the RC^-/AR^- case was not observed in the RC/AR case. This could be caused by the presence of intrinsic pause sites, which were identified during *in vitro* transcription assays using conditions of RC^-/AR^- (Figure 4.10 B). It is possible that in the RC/AR case, converging RNAP from P_Q knocks off elongating RNAP from P_X before reaching this locus.

The antisense effect (RC^-/AR) was restored (Figure 4.9 B, lane 4) by providing excess amounts of shorter 253 nt RNA from P_Q *in trans* from a high copy number plasmid to the construct used in lane 3. The 253 nt RNA species contained the sequence of the overlapping region and was used to represent Q_{AR} species in the mathematical model. Compared to the RC^-/AR^- case (Figure 4.9 B, lane 3), providing excess antisense Q_{AR} RNA reduced the expression of X and Anti-Q transcripts significantly. However, the extent of suppression of X (>700 nt) was lower than that observed in the RC/AR case, as the X transcript was still present even when large amounts of Q_{AR} RNA was supplied (Figure 4.10 D). Comparison of the intensity of X transcripts normalized to 5s RNA shown in lanes 3 (~0.9) and 4 (~0.3) (Figure 4.10 D) indicates that RNAP collision is responsible for two-thirds of the decrease of X transcription in the derepressed state, exerting a stronger effect on X transcription than antisense regulation. These results

suggest that both RNAP collision and antisense regulation are required to achieve high degree of suppression of X transcript upon derepression, although RNAP collision plays a more dominating role as predicted by the model (Figure 4.5 B). The presence of shorter RNA (100-223nt) in lane 4 compared to lane 3 suggests that these are products of endonucleolytic cleavage within paired regions formed by the "kissing" interactions between the loop structures of Q_{AR} and P_X transcripts (Figure 4.8) (Shokeen *et al.*, 2010).

Similarly, during convergent transcription Q_L was expressed at higher levels in the depressed state than repressed state (Figure 4.9 B, lane 5 vs 6), while higher abundance of truncated RNA was seen under repressed state (lane 6). In the absence of RNAP collision and antisense regulation, longer Q_L and Q_S RNA were expressed from P_Q, even under repressed state (lane 7, RC^-/AR^-). The lack of truncated RNA in this case in absence of RNAP collision (lanes 7) is consistent with our model that these RNAs are mainly a result of convergent transcription. When we provided excess amounts of antisense 104 nt Anti-Q RNA, representing X_{AR} species in the model, *in trans* from a high copy number plasmid (lane 8), Q_L expression decreased (lanes 7 vs 8, Figure 4.10 D), however it was visible at longer exposures. Importantly, Q_L RNA was expressed in RC^-/AR^- and RC^-/AR cases even under repressed conditions. Intensity of bands corresponding to both X (lane 1-4) and Q_L RNA in (lanes 5-8) normalized to 5s RNA, show that the high degree of suppression of X in the derepressed state (lane 1) and Q_L in the repressed state (lane 6) occurs only during convergent transcription (Figure 4.10 D). These results are consistent with model prediction.

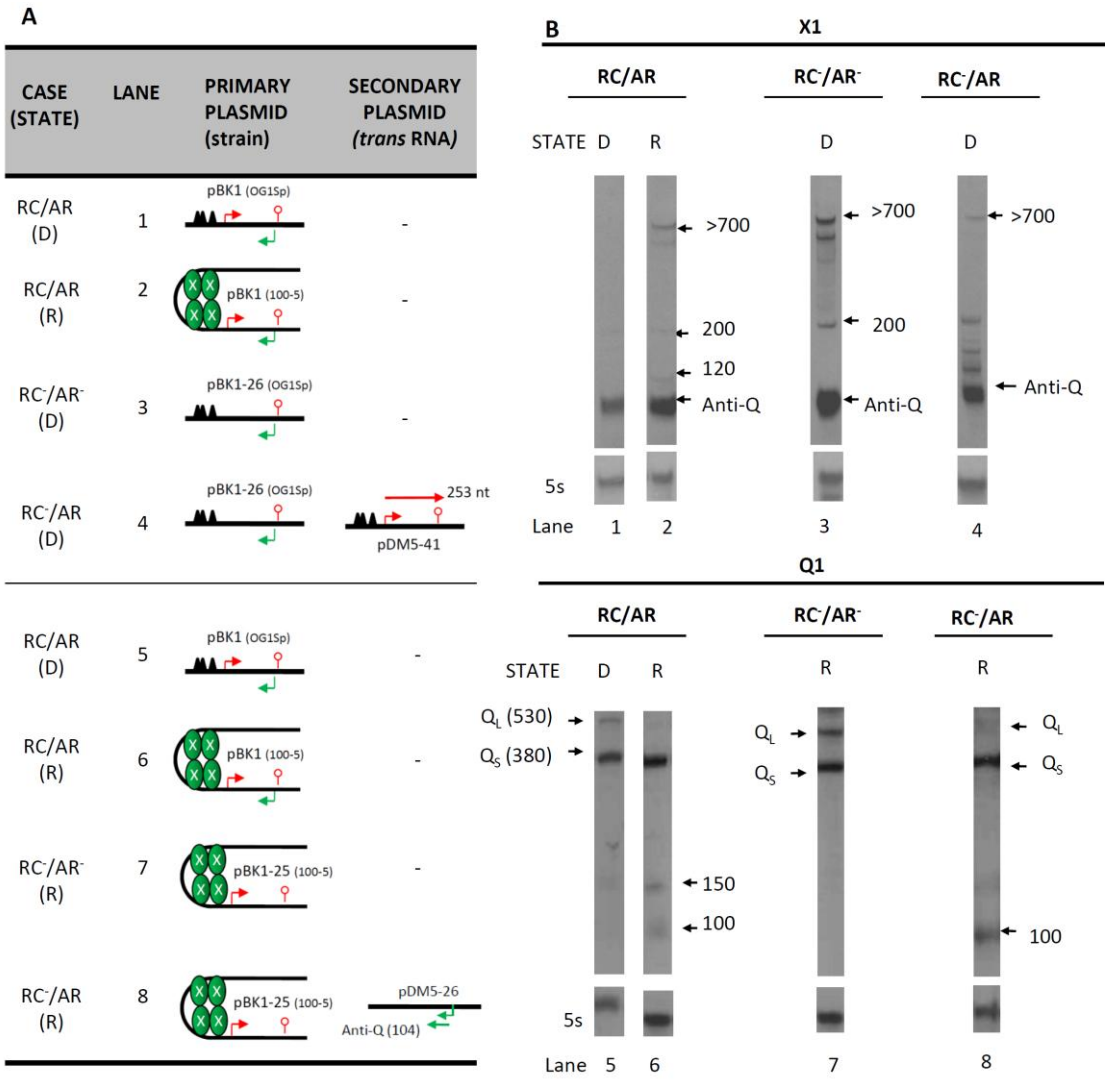


Figure 4.9: Inverse relationship of Q_L and X expression and presence of truncated RNAs in the *prgQ-prgX* locus. (A) Q_L and X RNA levels measured by qPCR using X-specific and Q_L-specific primers from OG1Sp (pCF10) under repressed (0 ng/mL cCF10) or derepressed (50 ng/mL cCF10) conditions. **(B)** Levels of Q_L RNA in response to different concentrations of cCF10. Data shown in (A) and (B) are averages of three independent experiments (error bars are standard deviations from mean values). **(C)** Top: Map for the single stranded RNA probes Q1, Q2, X1 and X2 used for Northern blotting experiments. Bottom: Northern blots to detect transcripts from P_Q and P_X in cells carrying WT pCF10, in either derepressed (D) or repressed (R) state. Probes/lanes: Q1/1-2, Q2/3-4, X1/5-6 and X2/7-8. **(D)** Quantitative estimation of intensity of bands for truncated P_Q and P_X transcripts in lanes 1-2 and 5-6 normalized to intensity of Q_S+Q_L and X respectively using ImageJ software (NIH).

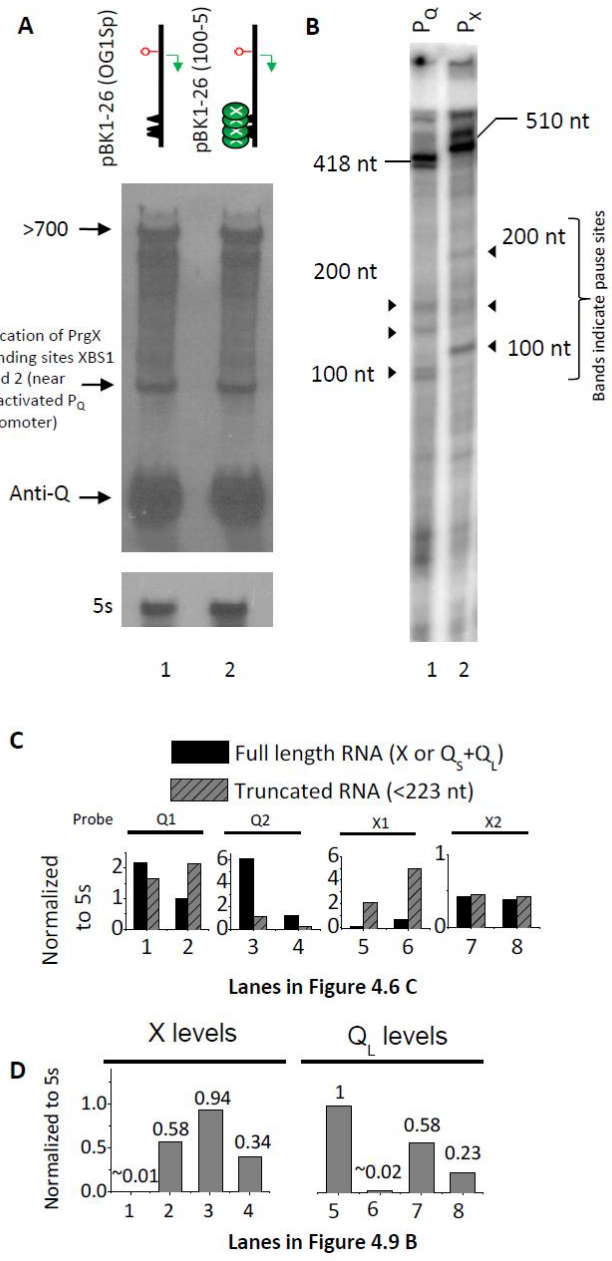


Figure 4.10: Roadblock, pausing and estimation of transcript levels. (A) PrgX binding to operators XBS1 and 2 does not pose as a roadblock for transcription from promoter P_X. Northern blot using X1 probe, showing similar expression pattern from single P_X promoter in the absence (lane 1) and presence (lane 2) of PrgX. PrgX binding sites XBS1 and 2, present near the inactivated promoter P_Q, and are indicated on the gel. **(B)** *In vitro* transcription showing presence of shorter terminated RNA from single promoter templates for P_Q (lane 1) and P_X (lane 2), indicating presence of pause sites. **(C)** Quantitative estimation of intensities of corresponding bands in Figure 4.6 C normalized to 5s RNA. **(D)** Quantitative estimation of intensity of bands corresponding to X transcript in Figure 4.9 B, lanes 1-4 normalized to intensity of 5s RNA and intensity of bands corresponding to Q_L transcript in Figure 4.9 B, lanes 5-8 normalized to intensity of 5s RNA, using ImageJ software (NIH). The band corresponding to X in the derepressed state (lane 1) and Q_L in the repressed state (lane 6) is rather faint in Figure 4.9 B, hence the intensities measured using Image J are only an approximation.

4.4 Discussion

Our results demonstrate that a head-to-head convergent promoter system can accomplish two levels of regulation: (1) enhancing the degree of repression through RNAP collision, and (2) suppressing transcript levels through interactions with *cis*-encoded antisense RNA produced from the overlapping region. With RNAP collision alone, an 8.3 fold increase in P_Q strength is predicted to give rise to a 12 fold increase in successful transcription from P_Q (Figure 4.5 A) relative to the repressed state. Experimentally a much greater difference in Q_L transcript level was observed using both a reporter system (Figure 4.2) and direct measurement of Q_L transcripts (Figure 4.6 A). This enhanced difference is accomplished by a greater degree of suppression of Q_L transcript level under the repressed conditions through the products of RNAP collision, namely truncated RNAs which have antisense regulatory effect on transcripts from both P_Q and P_X.

Using a combination of host strains and constructs to mimic conditions of RNAP collision (RC) and antisense regulation (AR), we show that Q_L transcript level which was suppressed under repressed conditions with RC/AR, became elevated when RC or both RC/AR were eliminated (Figure 4.9 B). Providing constant levels of Anti-Q RNA did not restore switch behavior, as demonstrated both experimentally (Figure 4.9 B, lane 8) and theoretically (Figure 4.5 D). The sharp difference in the level of Anti-Q/X_{AR} and Q species between repressed and derepressed state due to RC and AR contributes to the switch behavior.

Both mathematical modeling (Figure 4.5 A-B) and experimental observations (Figure 4.6 C), indicate the presence of certain prominent sizes of truncated RNA. However, the stochasticity of RNAP firing from promoters is expected to blur the discrete nature of truncated RNAs, pointing towards presence of potential RNAP pause sites within overlapping DNA, thus increasing the probability of collision at those sites. Such an effect will further strengthen the bistability due to enhanced effects of RNAP collision and antisense regulation (Palmer *et al.*, 2009).

Many biological switches have been shown to exhibit bistability in steady state analysis of their system models (Chatterjee *et al.*, 2008; Kobayashi *et al.*, 2004). The pivotal role of RC and AR in affording a bistable switch behavior is predicted by our model (Figure 4.5 C-D). Bistability allows for a binary behavior. Since conjugation is an energy intensive process, the donor cells must turn their conjugation system ON only in response to true signal. A switch type of regulatory mechanism allows for a tighter control of conjugation.

Transcription interference has been reported to exert suppressive effect on transcription from the weaker promoter in pR-pL and PR-PRE promoter pairs of bacteriophages 186 and λ respectively (Callen *et al.*, 2004; Palmer *et al.*, 2009). In these studies, RNAP collision was the major contributor to the observed interference whereas antisense effects were concluded to be insignificant. Convergent transcription in the *IME4* locus of *Saccharomyces cerevisiae*, was shown to result in antisense and sense transcripts which were inversely related between diploid and haploid cell-types (Hongay

et al., 2006). Again, in this case, the inverse relationship was ascribed to of transcriptional interference but not to antisense regulation. Other convergent transcription systems have also been reported to have inverse transcript levels under different conditions, including *ecsit* gene involved in mouse stem cell development (Dinger *et al.*, 2008), and the *mogR* locus of *L. monocytogenes* controlling cell motility (Guell *et al.*, 2009). In a study of transcription of *Pot2* locus in rice blast fungus truncated transcripts were observed that led to the identification of an antisense promoter (Kimura and Yamaguchi, 1998). *In silico* analysis on mouse genome and follow-up study using oligo-microarray revealed nearly two thousand sense-antisense transcripts (Kiyosawa *et al.*, 2005). The presence of multiple-sized transcripts without poly(A) appeared to hint the presence of truncated transcripts.

The fact that convergent transcription is ubiquitous and has persisted in evolution (Chen *et al.*, 2005) is perhaps an indication that such gene organizations confer fundamental mechanisms of gene regulation. We show that using a single repressor, convergent transcription allows for simultaneous regulation of two genes, giving rise to a complex, tightly regulated switch behavior. Critical to conferring the sophisticated control is the combined use of RNAP collision and antisense regulation. The collision frequency and the level of transcription suppression is highly dependent on many factors, including relative promoter strength, the length of overlap and the exact sequence that causes RNAP pausing and interaction. With such a wide range of possible outcomes,

using subtle structural tuning, convergent transcription may be highly adaptable to become robust controller for many complex cellular events.

4.5 Materials and Methods

4.5.1 Strains, reporter constructs, medium and growth conditions: *E.faecalis* strain OG1Sp resistant to Spectinomycin, and 100-5, a derivative of OG1Sp that constitutively produces the PrgX repressor from the chromosome were used for all experiments. *E.faecalis* cells were grown in THB (Difco) plates with 1.5% agarose (solid media) or M9-YE media supplemented with glucose (20%), 1mM $MgSO_4$ and 0.1mM $CaCl_2$ for liquid cultures. *E.faecalis* cultures grown overnight (ON) in M9-YE medium under selection pressure. ON cultures were diluted by 1:5 in fresh medium without antibiotics in presence (50 ng/mL) for derepressed culture or absence of cCF10 (0 ng/mL) for repressed culture, and grown for 90 min at 37°C under 110 rpm. Concentrations of antibiotics for *E.faecalis* were Spectinomycin, 100 μg/mL; Chloramphenicol, 20 μg/mL; Erythromycin, 10 μg/mL, and Tetracycline, 10 μg/mL. *E.coli* DH5α was used as a host for cloning the plasmids of interest. *E.coli* cells were grown on LB plates with 1.5% agarose (solid media) and LB broth and YENB broth (containing 0.075% Yeast Extract and 0.08% Nutrient Broth, Difco). Concentrations of antibiotics for *E.coli* were chloramphenicol, 20 μg/mL and Erythromycin, 200 μg/mL.

4.5.2 Reporter constructs: Plasmid pBK1 contains sequences from pCF10 (Dunny *et al.*, 1978) including the P_Q promoter through the IRS1 terminator, which also includes the P_X

promoter but does not include the *prgX* Orf and is described in (Johnson *et al.*, 2010). Plasmid pCJ1 was constructed by PCR amplification of a fragment from pBK1, which included pCF10 sequences between P_Q transcription start site (-1) through IRS1 (+390) with a primer encoded *Bgl*II site at the -1 end. This fragment was digested with *Bgl*II and *Bam*HI followed by ligation into a pBK1 backbone digested with *Xho*I and *Bam*HI. This reporter contains *lacZ* transcriptionally fused at P_X +221 and does not contain P_Q. To generate pBK1-26, two fragments from pBK1 were amplified, followed by digestion of the fragments using *Bsr*GI, ligation and amplification; followed by digestion with *Xho*I and *Bam*HI and ligation into a similarly digested pBK1 backbone. This reporter contains 2 substitutions within the P_Q -10 region that creates a *Bsr*GI site (Figure 4.4 A). To generate pBK1-25 we amplified a fragment from pBK1 using the primers containing a substitution in the P_X -10 region designed to inactivate the promoter. We digested the fragment with *Xho*I and *Xba*I and ligated it into a similarly digested pBK1 backbone. The mutated fragment from pBK1-26 was amplified to generate pDM5-26, followed by digestion with *Xho*I-*Xba*I and ligation it into a similarly digested pDM5 backbone, which contains Anti-Q region with the *prgX* terminator appended downstream of Anti-Q (Bae *et al.*, 2000). pDM5-41 was generated by cutting pDM5-25 (P_X mutant) with *Xba*I and *Kas*I, treating the backbone with Klenow fragment to generate blunt ends and self-ligating the backbone. This fused the right end of the pCF10 fragment at *prgQ* +253 to the plasmid backbone, deleting approximately 130 base pairs from the 5' terminus of Qs RNA.

4.5.3 Northern Blot Analysis: Approximately 2.5 mL of culture was added to 5 mL of Bacteria RNA protect reagent (Qiagen), followed by 5 min incubation. Samples were flash frozen in Ethanol and dry ice and stored at -80°C. Total RNA was obtained using the RNA Fastprep kit as per manufacturer's protocol. 1 μg of RNA from each sample was electrophoresed on 5% TBE-Urea gels (Bio-Rad Laboratories, BIORAD# 345-0086). RNA was then transferred onto positively charged Nylon membrane (Roche) by electrophoresis with 1X TBE. Membranes were cross-linked using Stratalinker (Stratagene). The blots were probed with respective dig-labeled single stranded RNA probes at 68°C ON, followed by washing and detection and detected using Dig-specific Fab fragments and CDP-Star (Roche). Blots were re-probed with Dig end-labeled DNA probe specific for 5S rRNA at 50°C ON. Blots were scanned using a HP Scanjet TWAIN 5100 and quantified using open source software NIH ImageJ. Lanes in Figure 4.6 C and 4.9 B using the same probe were respectively obtained from the same gel. In Figure 4.9 B, the brightness/contrast of lanes 1-4 were adjusted to 30%/10% and lanes 5-8 to 0%/10%.

4.5.4 Probe preparation for Northern Blots: Four directional single stranded RNA probes were designed to probe the first and second half of the 223 nt overlapping region of Qs/QL and X as shown in Figure 4D (top panel). Q1, Q2 probes complementary to regions of Qs/Q_L span from +1 nt to +108 nt and +122 nt to +220 nt respectively;

whereas X1, X2 probes complimentary to regions of X spanning from +4 to +98 and +113 to +223 respectively (Figure 4.4 A). Primers were designed to amplify those segments of DNA using pBK1 as template using the forward primer /reverse primer sequences shown below:

Probe	**Primer Sequence (Forward primer/ Reverse primer)**
Q1	CGGATTTAGGTGACACTATAGAAGTAACTAATGTTGCAACAAA CG/ GAACCGACTGCCATAGGACG
Q2	CGGATTTAGGTGACACTATAGAATTGTTACTATGATTTTTTCTA TCAG/TAGGAGGGGTGTAAATGAAAAC
X1	CGGATTTAGGTGACACTATAGAACCGACTGCCATAGGACG/ GTAACTAATGTTGCAACAAACG
X2	CGGATTTAGGTGACACTATAGAATAGGAGGGGTGTAAATGAAA AC /AATTGTTACTATGATTTTTTCTATCAG

An SP6 promoter with sequence CGGATTTAGGTGACACTATAGAA was inserted in front of the PCR amplified DNA segments and used as for *in vitro* transcription. SP6 polymerase and digoxygenin-labeled UTP (Roche) were used to generate dig-UTP labeled RNA probes.

4.5.5 *Quantitative PCR*: 0.6 mL of culture was added to 1.2 mL of Bacteria RNA protect reagent (Qiagen), followed by 5 min incubation. Samples were flash frozen in Ethanol and dry ice and stored at -80°C. Cells were lysed with lysozyme (30 mg/mL) in TE buffer + Mutanolysin (500 U/mL). RNA was prepared using RNeasy kit (Qiagen). 5 μg of RNA was subjected to rigorous DNAse treatment with Turbo DNAse (Ambion) as per manufacturer's instructions, followed by reverse transcription of 100 ng of DNA free

RNA using Superscript III first strand synthesis kit (Invitrogen 18080-051) with random primers. Approximately 1.5 ng of cDNA was used for quantitative PCR using Brilliant SYBR Green QPCR Master Mix (Stratagene, Agilent Technologies) using gene specific primers on real-time MX3000P instrument (Stratagene, La Jolla, CA). Transcript levels were determined using the ΔCt method with respect to *GyraseB*, a moderately expressed housekeeping gene used as a reference.

Primers used for qPCR

Transcript	**Primer Sequence (Forward primer/ Reverse primer)**
Q_L	GCTTTGGCTCAATCCTTTG / CTGACTCTCGACCGATTTCTG
X	AAAATGACCCGCTTGTGG / GTTGTTTGCACGGCTCTTAC
lacZ	CGCTTTAATGATGATTTCAGCC/ TGCCATAAAGAAACTGTTACCC

4.5.6 Induction of pCF10 (OG1Sp) cells: Overnight cultures of single colony of OG1Sp cells containing pCF10 plasmid were grown in complete M9-YE with 10 μg/mL Tetracycline. The overnight culture was split into 5 tubes and diluted 1:10 in complete M9-YE without antibiotics containing five different concentrations of cCF10 (0, 0.05, 0.5, 5 and 50 ng/mL). The cells were incubated with the various cCF10 concentrations for 60 minutes, after which cells were harvested for RNA. After RNA extraction, DNAse treatment and cDNA synthesis, qPCR was performed to measure the Q_L expression relative to *Gyrase B*.

4.5.7 Switching ON and OFF response with pBK1 (100-5) cells: Overnight culture of single colony of strain 100-5 containing pBK1plasmid was grown in complete M9-YE with 20 μg/mL Chloramphenicol. The overnight culture was split into 4 tubes and diluted

1:10 in complete M9-YE without antibiotics containing four different concentrations of cCF10 corresponding to OFF (0, 0.05 ng/mL) and ON (5 and 50 ng/mL) response based on Figure 4.6 B. The culture was supplemented with 5 ng/mL of iCF10. To obtain the turning ON response, cells were incubated with the various cCF10 concentrations for 60 minutes, after which cells were harvested for RNA. To obtain the turning OFF response, cells were exposed to 50 ng/mL of cCF10 for 60 min. Equal volumes of this induced culture were washed 1X with fresh M9-YE medium. Cell pellets were collected via centrifugation followed by removal of supernatant. Cell pellets were re-suspended in equal volume of M9-YE medium containing different concentrations of cCF10 (0, 0.05, 5 and 50 ng/mL). After 60 minutes of incubation, equal volumes of cultures were harvested for RNA. After RNA extraction, DNAse treatment and cDNA synthesis, qPCR was performed to measure the lacZ expression relative to Gyrase B.

4.5.8 Protein preparation: Overnight culture (16 h) of strain OG1Sp carrying pCF10 plasmid and strain 100-5 were grown in M9-YE at 37ºC with antibiotics Spectinomycin (100 μg/ml) and Tetracycline (10 μg/ml). The overnight culture was diluted 1:10 in fresh M9-YE and grown for 2.5 h at 37ºC. Equivalent amount of cells were collected by centrifugation. The cells were re-suspended in 50 μl TES solution (10 mm tris-hcl, ph 8.0, 10 mM EDTA, 25% sucrose) and aliquoted for total cell counts. Equal volume of TES containing 30 mg/mL Lysozyme (sigma) was added to cells and incubated at 37ºC for 15 min. Protoplasts were re-suspended in 25-50 μL 1X sample buffer (2% SDS, 0.625

mm Tris-HCl, pH 6.8, 10% glycerol, 2.5% β-mercaptoethanol, 0.005% bromphenol blue) for SDS-page, boiled for 3 min. and centrifuged 6000 rpm, for 3 min.

4.5.9 Western blotting: Purified His-Prgx and cell lysates were resolved by denaturing electrophoresis on 12% polyacrylamide gels (SDS-page) and transferred onto 0.2 μm nitrocellulose membrane (S&S) using Towbin buffer at 90 Volts for 70 minutes at 4ºC. PrgX was detected by using the His-PrgX polyclonal antibody (Bae et al., 2000) at a dilution of 1:1500 and Horseradish peroxidase (Hrp)-Goat anti-rabbit IGG (Zymed) at 1:5000. Detection was performed using super signal west pico chemiluminescent substrate according to manufacture (Thermo scientific).

4.5.10 Flow cytometry: The steady state expression levels of the GFP protein was measured from convergent promoter construct pBK2ΔlacZ::gfpmut3b which has gfp reporter fused downstream of IRS1 (Figure 4.10 E) in *E.faecalis* cells strain OG1RF (Kozlowicz, 2005; Kristich et al., 2007). Cells were grown overnight in M9-YE media supplemented with glucose (20%), 1mM MgSO4 and 0.1mM CaCl2 and Chloramphenicol (20 μg/mL). Overnight cultures were diluted 1:20 into fresh M9 medium and grown for 90 minutes at 37°C. The cells were then induced with different concentrations of cCF10 in M9 medium without antibiotic selection at 37°C for 60 minutes after which steady state was reached. Cell suspensions were analyzed via flow cytometry using a FACS Calibur flow cytometer (BD Biosciences). Live cells were gated

based on cell size. Expression of GFP was measured for 100,000-200,000 live cells using the FL-1 laser with a wavelength of 488nm. Data was analyzed using FlowJo Flow Cytometry Analysis Software.

4.5.10 In vitro transcription assay: A linear DNA template of length 705 bp, containing single promoter P_Q was generated by PCR amplification of the region between -287 to +418 relative to *prgQ* transcription start site present on plasmid pBK1-25 (Figure 4.10 A) using forward primer F1 (CATCCAGTCTATTAATTGTTGCCGG) and reverse primer R1 (TAGTTTTCTCCTCCTTTTGGATCC) to give a 418 nt run off product from P_Q. Similarly, a linear DNA template of length 705 bp containing single promoter P_X was generated by performing PCR on plasmid pBK1-26 (-287 to +418 relative to *prgQ* transcription start site, Figure 5A) using forward primer F1 and reverse primer R1, to give a 510 nt run off product from P_X. The resulting P_Q template or P_X templates (2.5 nM) was mixed with *Bacillus subtilis* RNAP ((Johnson *et al.*, 2010) and references therein) in a 35μl reaction mixture containing 20 mM Tris-HCl, pH 7.9; 20 mM NaCl; 20 mM $MgCl_2$; 0.1 mM EDTA; 11 mM β-mercaptoethanol; 286 μM each of ATP, GTP, and CTP; 9.53 μM UTP; 2.5 μg of the initiating dinucleotide ApU and GpU (Ribomed); and 6 μCi [32P] UTP (MP Biomedicals). Reaction mixtures were incubated at 37°C for 15 min and stopped by adding an equal volume of isoamyl alcohol: phenol:chloroform, pH 5.2. RNA fragments were obtained by recovering the aqueous phase. The RNA was resolved on a 13.5% denaturing polyacrylamide gel, and scanned on a phosphorimager (Molecular Dynamics). Gels were analyzed using Image Quant software.

CHAPTER 5

MATHEMATICAL MODEL OF PRGQ-PRGX GENE NETWORK: COUPLED EFFECT OF RNA POLYMERASE COLLISION AND ANTISENSE INTERACTION

5.1 Summary

This chapter discusses the mathematical model based on the *prgX/prgQ* genetic network of plasmid pCF10 in *Enterococcus faecalis*. This chapter is a detailed analysis of the mathematical model described in Chapter 4. The assumptions and construction of the discrete mathematical model for estimating effect of transcriptional interference in the *prgX/prgQ* locus is described. The model parameters estimated from the transcriptional interference model are used for the ODE based model, which is then used for steady state analysis of the *prgX/prgQ* system. Non-dimensional parameter analysis is used for evaluating the robustness of predicted steady state behavior over a range of parameter space.

5.2 Discrete mathematical model of transcriptional interference due to RNA polymerase collision

RNA polymerase (RNAP) collision in the overlapping region between P_X and P_Q driving expression of respective *prgX* and *prgQ* genes results in failed transcription (Shearwin *et al.,* 2005). The frequency of collision, and thus its effect on apparent transcription rate is influenced by the traveling time of RNAP in this overlapping region and the transcription initiation frequency. We model this process using a discrete mathematical model that estimates the frequency of collision of RNAP in this locus as well as an averaged transcription rate for both full length (Q and X) and truncated RNA (Q_{AR} and X_{AR}) generated from this locus. These rates are used in the subsequent deterministic ODE based model for steady-state analysis. Since the P_Q-P_X promoters are separated by sufficient length of 223 bp, RNAP is assumed to bind independently at either promoter. Transcriptional Interference as a result of collision between two RNAPs whether both elongating or one is stationary RNAP (Figure 4.5 A-B) (Shearwin *et al.*, 2005) is considered, in addition to promoter occlusion due to passage of RNAP originating from the opposing promoter. It is assumed that abundant RNAPs are available for binding to P_X and P_Q (Sotiropoulos and Kaznessis, 2007). The frequency of RNAP binding to promoters P_Q and P_X is denoted by f_Q and f_X respectively, and the time taken to form the elongation complex at respective promoters is denoted by t_{iQ} and t_{iX} (Figure 5.1). The model has the following rules and assumptions:

1. 80% of the collisions between RNAPs, whether moving or stationary, result in both RNAP dropping off the DNA, resulting in failure of transcription (Crampton *et al.*, 2006). The remaining 20% of collided RNAP complexes, one of the RNAP is allowed to continue to elongate, while the other is considered to abort transcription. This gives rise to an overall 90% RNAP fall-off rate (80% +10%) and 10% survival rate for each round of transcription from both promoters. This is based on a 10% level of X transcript observed upon derepression (Figure 4.6 A). *In vitro* studies also suggest non-zero survival rates upon RNAP collision (Crampton *et al.*, 2006). Under the derepressed state an 8.3 fold stronger P_Q (Figure 4.2) would have resulted in almost complete suppression of X transcription. The results in Figure 4.6 A suggest approximately 10% of the transcription events from P_X succeed. Adopting a 90% RNAP falling-off rate upon collision in our simulation shows a 7 fold reduction of X transcript under derepressed condition as compared to repressed condition (Figure 4.5 B), consistent with 9 fold reduction observed in experiments (Figure 4.6 A).
2. As a result RNAP collision, truncated RNA from P_Q and P_X capable of antisense interactions are generated. These are denoted as Q_{AR} and X_{AR} with sizes between 180-223 nt and 104-223 nt respectively (Figure 5.1). This is based on *in silico* structural analysis of truncated RNA from this locus (Figure 4.8) (Shokeen *et al.*, 2010).

3. An elongating RNAP is not hindered by binding of PrgX protein to the DNA *(*Figure 5.5A).
4. Once an elongating RNAP has traversed the overlapping DNA, a successful transcript would be transcribed. If the elongating RNAP originates from P_Q, then nascent Q RNA will be transcribed, else if RNAP originated from P_X, X would be formed.
5. RNAP moves at a velocity of elongation at 50 bp/s, which has been reported as the average velocity of RNAP in *E.coli* in presence of pausing (Neuman *et al.*, 2003).

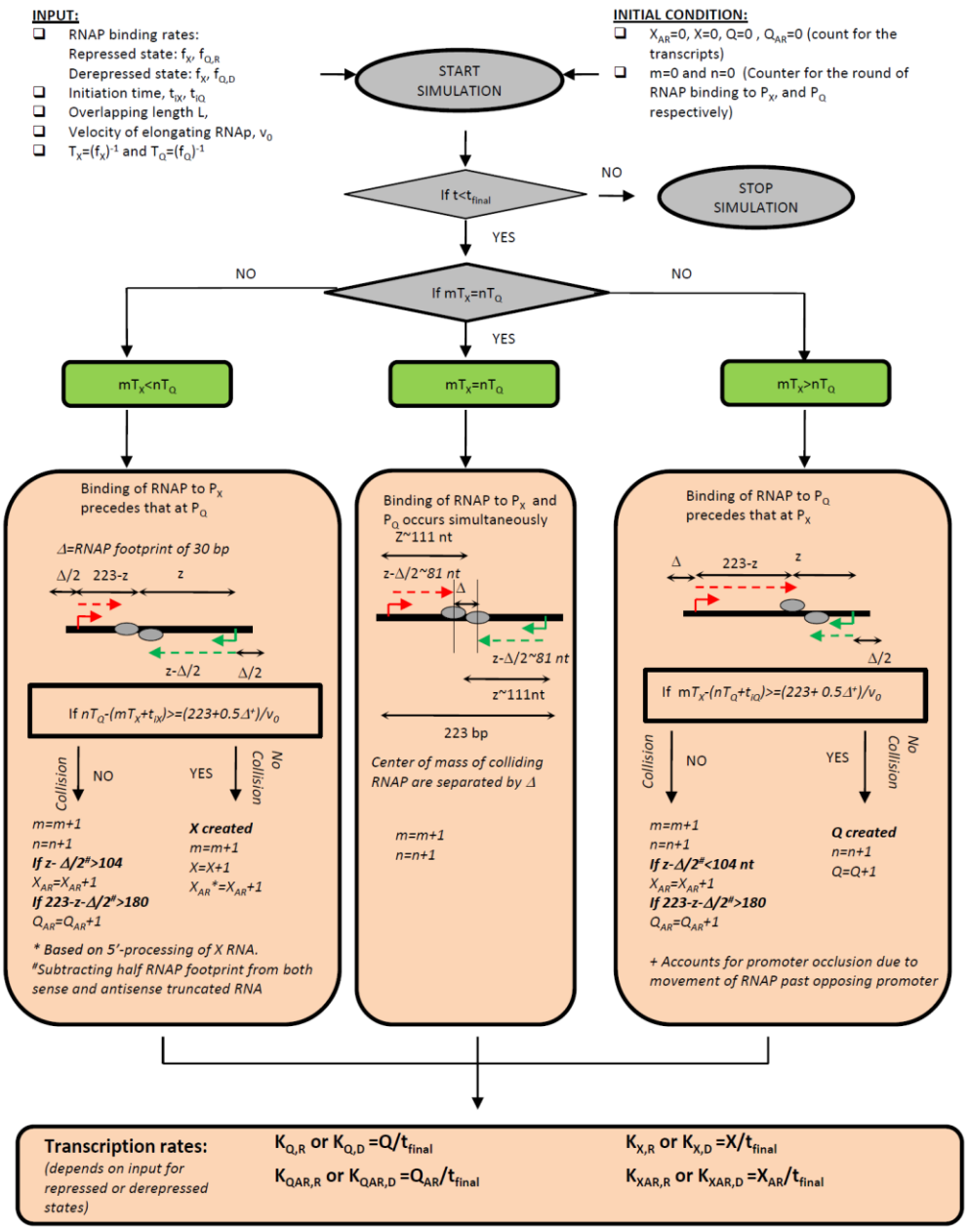


Figure 5.1: Algorithm used to determine loci and frequency of collision within overlapping region of *prgQ* and *prgX* for the discrete model. Q_{AR} and X_{AR} denote the sub-population of truncated RNA within sizes 180-223 nt and 104-223 nt generated from P_Q and P_X promoters respectively. The transcription rates of Q, X, Q_{AR} and X_{AR} for the RC/AR case are calculated.

In order to simulate the P_Q and P_X expression under derepressed (denoted by subscript D) and repressed conditions (denoted by subscript R), the intrinsic RNAP binding frequency at P_Q, $f_{Q,D}$ was set at 0.1/s under derepressed state based on RNAP initiation rates estimated for P_{bla} system in *E.coli* (Liang *et al.*, 1999) and PR-PRE system in bacteriophage λ (Palmer *et al.*, 2009). Under repressed state P_Q activity decreases by 9 fold (Figure 4.2), thus RNAP binding frequency under repressed conditions $f_{Q,R}$ was set to 0.011/s (ratio of 0.1/9).The binding frequency of RNAP at the constitutive P_X promoter, f_X, was set at 0.012/s (ratio of 0.1/8.3) under both derepressed and repressed conditions, while t_{iQ} and t_{iX}, time delay for formation of elongation complex from closed complexes at P_Q and P_X respectively, were set at 10 s for all conditions. This is based on real-time observation of transition from initiation complex to elongation complex for T7 RNAP and T7 RNAP consensus promoter using single-molecule and stopped-flow FRET methods (Tang *et al.*, 2009). These were used as the input to the discrete model as shown in Figure 5.1.

At time t=0, RNAPs occupy both promoters. Initially, n=0 and m=0, where n and m are the counter for number of rounds of transcription from P_Q and P_X respectively (Figure 5.1). The first round of transcription, n=1 and m=1, inevitably results in RNAP collision at the middle of the overlapping region at L/2~111 bp (RNAP pausing at specific sites is not considered, instead an average RNAP velocity of 50 bp/s is considered). Considering the RNAP footprint, the truncated RNA generated would have length of 81 nt (L/2-30 nt). Thereafter, depending on RNAP firing rates from P_Q and P_X, the model evolves to

generate successful or truncated transcripts. The movement of RNAP along the DNA is simulated in time, the n^{th} and m^{th} round of transcription from P_Q and P_X are recorded. For each round of successful transcription, an elongating RNAP from P_Q has to travel 223 bp (overlapping DNA)+ 30 bp (Half of RNAP footprint) without encountering an RNAP from P_X, and vice versa. If during the time RNAP from P_Q traverses P_X, another RNAP molecule binds at P_X, such an event is considered to result in occlusion of P_X due to RNAP collision. Under this condition, the value of m is incremented, i.e. RNAP is considered to be knocked off of P_X. Similarly, occlusion of P_Q due to movement of RNAP from P_X is also considered.

For each successful round of transcription from P_Q and P_X the value of Q and X are incremented respectively. For each successful X RNA, the 104 nt transcript from P_X, X_{AR} is incremented, based on 5' end processing of X RNA (Bae *et al.*, 2000). Every event of collision in the overlapping region and length of truncated RNA generated from both P_Q and P_X are recorded. Simulations were performed for a minimum of 1000 rounds of transcription for each parameter set corresponding to the repressed and derepressed state of promoter P_Q. The apparent transcription rate for full length transcripts Q, X and truncated transcripts Q_{AR} and X_{AR} was calculated by dividing the cumulative amount of RNA estimated at the end of simulation by the time of simulation (Figure 5.1). While Q_{AR} species considers truncated P_Q RNA within 180-223 nt only, X_{AR} includes truncated RNA and Anti-Q RNA generated between 104-223 nt originating at P_X. The algorithm used is summarized in Figure 5.1.

Table 5.1 List of variables and parameters used in the mathematical model

Notation	Description	Notation	Description
O	DNA of plasmid in loop form	C	Intracellular cCF10
N	Plasmid copy number, equal to 5	$PrgX$	Repressor PrgX
Q_S	Q_S mRNA	B	DNA binding sites XBS 1 and 2
Q_L	Q_L mRNA	$(PrgX)_2$	PrgX dimer
X_{AR}	Truncated P_X RNA with sizes between 104-223 nt	$(PrgX)_2I_2$	iCF10 bound PrgX dimer
X	X mRNA	$(PrgX)_2C_2$	cCF10 bound PrgX dimer
Q_{AR}	Truncated P_Q RNA with sizes between 180-223 nt	$(PrgX)_4B$	PrgX tetramer bound to XBS 1 and 2 causing DNA looping (repressed state)
i	Extracellular iCF10	$(PrgX)_4I_4B$	iCF10 bound PrgX tetramer bound to XBS 1 and 2 causing DNA looping (repressed state)
I	Intracellular iCF10	$((PrgX)_2C_2)_2B$	cCF10 bound PrgX dimers bound to XBS 1 and 2 resulting in linear DNA (derepressed state)

Parameter	Description	Value used for bistability	Units
μ	Specific growth rate of donor cells	2.58×10^{-4}	s^{-1}
$f_{Q,R}$	RNAP binding rate at P_Q under repressed state	1.1×10^{-2}	s^{-1}
$f_{Q,D}$	RNAP binding rate at P_Q under derepressed state	1.0×10^{-1}	s^{-1}
f_X	RNAP binding rate from constitutive promoter P_X	1.2×10^{-2}	s^{-1}
$K_{Q,R}$	Transcription rate constant of Q in repressed state	7.23×10^{-3}	s^{-1}
$K_{Q,D}$	Transcription rate constant of Q in derepressed state	8.87×10^{-2}	s^{-1}
$K_{X_{AR},R}$	Transcription rate constant of X_{AR} in repressed state	1.02×10^{-2}	s^{-1}
$K_{X_{AR},D}$	Transcription rate constant of X_{AR} in derepressed state	1.21×10^{-3}	s^{-1}
$K_{X,R}$	Transcription rate constant of X in repressed state	8.23×10^{-3}	s^{-1}
$K_{X,D}$	Transcription rate constant of X in derepressed state	1.21×10^{-3}	s^{-1}
$K_{Q_{AR},R}$	Transcription rate constant of Q_{AR} in repressed state	1.80×10^{-3}	s^{-1}
$K_{Q_{AR},D}$	Transcription rate constant of Q_{AR} in derepressed state	1.08×10^{-2}	s^{-1}
K_i	Generation rate of extracellular iCF10	1×10^{-3}	s^{-1}
V_{conv}	Volume conversion factor	1	-
K_{T_i}	Transport rate constant of iCF10	1×10^{-3}	s^{-1}
K_{T_C}	Transport rate constant of cCF10	1×10^{-3}	s^{-1}
K_{AR}	Rate constant of interaction between X and Q_{AR} RNA	1×10^{-3}	$(nM)^{-1}$. s
$K_{Q-X_{AR}}$	Equilibrium constant of Q and X_{AR} interaction	4.43	$(nM)^{-1}$
K_B	Equilibrium constant of DNA binding reaction	1×10^{4}	-
λ_{Q_S}	Degradation rate of Q_S mRNA	1×10^{-3}	s^{-1}
λ_{Q_L}	Degradation rate of Q_L mRNA	1×10^{-1}	s^{-1}
$\lambda_{X_{AR}}$	Degradation rate of X_{AR} RNA	1.36×10^{-4}	s^{-1}
$\lambda_{Q_{AR}}$	Degradation rate of Q_{AR} RNA	1×10^{-3}	s^{-1}
λ_X	Degradation rate of X RNA	1×10^{-4}	s^{-1}

λ_i	Degradation rate of extracellular iCF10	1×10^{-6}	s^{-1}
λ_I	Degradation rate of intracellular iCF10	1×10^{-6}	s^{-1}
λ_C	Degradation rate of intracellular cCF10	1×10^{-6}	s^{-1}

The rate constants for transcription under repressed /derepressed state after considering RNAP collision effects are summarized in Table 5.1, where transcription of full length Q RNA under repressed and derepressed state is respectively denoted by $K_{Q,R}$ and $K_{Q,D}$, full length X RNA by $K_{X,R}$ and $K_{X,D}$, Q_{AR} RNA by $K_{Q_{AR},R}$ and $K_{Q_{AR},D}$ and X_{AR} RNA by $K_{X_{AR},R}$ and $K_{X_{AR},D}$.

5.3 Transcription rates from P_Q and P_X under repressed and derepressed conditions

The derepressed or repressed state of P_Q is determined by the formation of DNA loop in the region of operator sites XBS 1 and XBS 2 (Figure 4.1). Binding of iCF10 to PrgX causes PrgX-iCF10 tetramer to occupy operator sites XBS 1 and 2 and form a looped structure resulting in P_Q repression (Figure 4.1 B). Conversely, cCF10 binding to PrgX causes PrgX-cCF10 dimer formation that results in the operator sites to return to linear form, thus relieving P_Q repression (Figure 4.1 C). The binding of PrgX-iCF10 and PrgX-cCF10 dimers to XBS 1 and 2 is treated as a rapid equilibrium process (Shi *et al.*, 2005). The equilibrium binding reactions describing the formation of PrgX dimer ($(\Pr gX)_2$), iCF10 bound PrgX dimer ($(\Pr gX)_2 I_2$), cCF10 bound PrgX dimer ($(\Pr gX)_2 C_2$), with equilibrium rate constants K_1, K_2 and K_3 respectively are shown in Equations 5.1-5.3.

The formation of DNA bound PrgX tetramers ($(\Pr gX)_4 B$), DNA bound PrgX-iCF10 tetramers ($(\Pr gX)_4 I_4 B$) and DNA bound PrgX-cCF10 dimers ($(\Pr gX)_2 C_{2\ 2} B$) are described in Equations 5.4-5.6, where K_{DNA} is the equilibrium rate constant describing binding to the DNA at operator sites XBS 1 and 2 (Shi *et al.*, 2005). The notation of all species involved is shown in Table 5.1.

$$2\Pr gX \xrightleftharpoons{K_1} (\Pr gX)_2 \qquad (5.1)$$

$$(\Pr gX)_2 + 2I \xrightleftharpoons{K_2} (\Pr gX)_2 I_2 \qquad (5.2)$$

$$(\Pr gX)_2 + 2C \xrightleftharpoons{K_3} (\Pr gX)_2 C_2 \qquad (5.3)$$

$$2\Pr gX_2 + B \xrightleftharpoons{K_{DNA}} (\Pr gX)_4 B \qquad (5.4)$$

$$2((\Pr gX)_2 I_2) + B \xrightleftharpoons{K_{DNA}} (\Pr gX)_4 I_4 B \qquad (5.5)$$

$$2((\Pr gX)_2 C_2) + B \xrightleftharpoons{K_{DNA}} ((\Pr gX)_2 C_2) B \qquad (5.6)$$

Applying rapid equilibrium assumption we obtain Equations 5.7-5.9.

$$\left[\ \Pr gX_{\ 4} B\right] = K_{DNA} K_1^{\ 2} [\Pr gX]^4 \cdot [B] \qquad (5.7)$$

$$\left[\ \Pr gX_{\ 4} I_4 B\right] = K_{DNA} K_2^{\ 2} K_1^{\ 2} [\Pr gX]^4 \cdot [I]^4 \cdot [B] \qquad (5.8)$$

$$\left[\ (\Pr gX)_2 C_{2\ 2} B\right] = K_{DNA} K_3^{\ 2} K_1^{\ 2} [\Pr gX]^4 \cdot [C]^4 \cdot [B] \qquad (5.9)$$

Being peptide molecules iCF10 and cCF10 concentrations are assumed to be higher than repressor PrgX. The value of binding constant allows us to assume PrgX bound to the DNA exists either as PrgX-iCF10 tetramer or two PrgX-cCF10 dimers. Thus $(\Pr gX)_4 B$ is negligible. The net sum of looped and unlooped XBS 1 and 2 sites would equal the copy number of the plasmid as shown in Equation 5.10 where [N] is the concentration of pCF10 plasmid in the donor cell. A plasmid copy number of 5 was used in simulations. The volume of the cell was assumed to be similar to *E.coli* at 1×10^{-15} liter (Sotiropoulos

and Kaznessis, 2007), such that 1 molecule inside the cell is equivalent to a concentration of 1.66 nM.

$$[B]+[(\Pr gX)_4 I_4 B]+[\ (\Pr gX)_2 C_{2\ \ 2} B]=[N] \qquad (5.10)$$

Since the number of PrgX molecules inside the cell is assumed to be large (Nakayama *et al.*, 1994), we assume that XBS 1 and 2 are mostly in an looped state, hence we neglect the term [B] in Equation 5.10. Substituting Equations 5.7-5.9 in 5.10 gives Equation 5.11.

$$[B]=\frac{N}{K_{DNA}K_2{}^2K_1{}^2\ [\Pr gX]^4\cdot[I]^4+\ K_{DNA}K_3{}^2K_1{}^2\ [\Pr gX]^4\cdot[C]^4} \qquad (5.11)$$

The concentration of pCF10 in the looped state ([O]), determined by $(\Pr gX)_4 I_4 B$ species, is obtained by substituting Equation 5.11 in 5.8 as shown in Equation 5.12.

$$DNAloop=[O]=\frac{[N][I]^4}{[I]^4+K_B[C]^4} \qquad (5.12)$$

We define the parameter, K_B, where $K_B=\left(\frac{K_3}{K_2}\right)^2$. Since at equilibrium $K_3\propto\frac{1}{[C]^2}$ and $K_2\propto\frac{1}{[I]^2}$ (based on Equations 5.8 and 5.9), considering equi-molar quantities of $(\Pr gX)_2 I_2$ and $(\Pr gX)_2 C_2$, the ratio of $K_B=\left(\frac{K_3}{K_2}\right)^2=\left(\frac{[I]}{[C]}\right)^4$. High performance liquid chromatography fractionation of supernatants of pCF10 carrying *E.faecalis* cultures show that iCF10 and cCF10 are secreted typically in a molar ratio of 10-100 to 1, at this ratio iCF10 is considered sufficient to neutralize the effect of cCF10 (Nakayama *et al.*,

1994). For a conservative estimate we set the ratio $\frac{[I]}{[C]}$ to 10, which gives an estimate of $K_B=10^4$.

The transcription initiation rate of from P_Q depends on whether P_Q promoter is in the derepressed state (unlooped state) or repressed state (looped state), whereas that for P_X, being constitutive, is considered constant. However, due to transcriptional interference exerted by P_Q and P_X on each other, the transcription from both P_Q and P_X will depend on the state of the pCF10 DNA, and hence on the number of looped and unlooped operator sites. The net transcription rate of full length and truncated RNA from P_Q and P_X is thus denoted by the sum of transcription rate in the repressed and derepressed states (summarized in Table 5.1) weighted by concentration of looped [O] to unlooped operator sites ([N]-[O]) respectively as shown in Equations 5.13-5.16.

$$k_Q = K_{Q,R}[O] + K_{Q,D}([N]-[O]) \quad (5.13)$$

$$k_X = K_{X,R}[O] + K_{X,D}([N]-[O]) \quad (5.14)$$

$$k_{Q_{AR}} = K_{Q_{AR},R}[O] + K_{Q_{AR},D}([N]-[O]) \quad (5.15)$$

$$k_{X_{AR}} = K_{X_{AR},R}[O] + K_{X_{AR},D}([N]-[O]) \quad (5.16)$$

5.4 Interactions of X_{AR} with full-length Q RNA and Q_{AR} with full-length X RNA

Transcription from P_X and P_Q generates both full length X, Q transcripts as well as truncated transcripts. Previous experimental data indicated that interaction of a 104 nt Anti-Q RNA (or X_{AR} in the RC/AR model) expressed from P_X promoter occurs between

stem loops I and II of Anti-Q and nascent Q transcript as shown in Figure 4.3 and Figure 4.8.

Interaction of Anti-Q with an elongating Q RNA causes a shift from an anti-terminator to terminator structure of Q RNA. This structural change results in termination of transcription of Q RNA (Johnson *et al.*, 2010) to give the shorter Q_S RNA. Since the stem loops I and II are made within the first 104 nt from P_X, we assume truncated RNA from P_X with length greater than 104 nt are capable of forming secondary structures resembling stem loops I and II that can interact with nascent Q RNA. As a result X_{AR} RNA species, which includes Anti-Q as well as truncated RNA form P_X between sizes 104-223 nt are considered to interact with Q RNA following irreversible second order kinetics with experimentally determined rate constant K_{AR} as 1×10^{-6} $M^{-1}s^{-1}$ (Shokeen *et al.*, 2010).

Table 5.2 List of dimensionless variables and parameters used in the ODE model

Species	Dimensionless form	Species	Dimensionless form
x_1	$\frac{[Q_s]}{[Q_s]_0}$	x_5	$\frac{[Q_{AR}]}{[Q_{AR}]_0}$
x_2	$\frac{[Q_L]}{[Q_L]_0}$	x_6	$\frac{[i]}{[i]_o}$
x_3	$\frac{[X_{AR}]}{[X_{AR}]_0}$	x_7	$\frac{I}{I_o}$
x_4	$\frac{[X]}{[X]_o}$	x_8	$\frac{[C]}{[C]_o}$

Dimensionless parameter (DP)	Dimensional form	DP	Dimensionless form	DP	Dimensional form
τ	$\mu \cdot t$	$k_{Q_{AR},D}$	$\frac{K_{Q_{AR},D}}{K_{Q_{AR},R}}$	λ_{x_2}	$\frac{\lambda_{Q_L}}{\mu}$
α	$\frac{O}{N}$	k_i	$\frac{K_i V_{conv}}{\mu}\frac{[Q_S]_0}{[i]_o}$	λ_{x_3}	$\frac{\lambda_{X_{AR}}}{\mu}$
$k_{Q,R}$	$N \cdot \frac{K_{Q,R}}{\mu \cdot Q_{S\ 0}}$	$k_{V_{conv}}$	$\frac{[i]_o}{[I]_o}\frac{1}{V_{conv}}$	λ_{x_4}	$\frac{\lambda_X}{\mu}$
$k_{Q,D}$	$\frac{K_{Q,D}}{K_{Q,R}}$	k_{T_i}	$\frac{K_{T_i} V_{conv}}{\mu}$	λ_{x_5}	$\frac{\lambda_{Q_{AR}}}{\mu}$
$k_{X_{AR},R}$	$N \cdot \frac{K_{X_{AR},R}}{\mu \cdot [X_{AR}]_0}$	k_{T_c}	$\frac{K_{T_c} V_{conv}}{\mu}$	λ_{x_6}	$\frac{\lambda_i}{\mu}$
$k_{X_{AR},D}$	$\frac{K_{X_{AR},D}}{K_{X_{AR},R}}$	k_{AR}	$\frac{K_{AR}}{\mu} \cdot [Q_{AR}]_o$	λ_{x_7}	$\frac{\lambda_I}{\mu}$
$k_{X,R}$	$N \cdot \frac{K_{X,R}}{\mu \cdot [X]_o}$	$k_{Q-X_{AR}}$	$K_{Q-X_{AR}} \cdot [X_{AR}]_0$	λ_{x_8}	$\frac{\lambda_C}{\mu}$
$k_{X,D}$	$\frac{K_{X,D}}{K_{X,R}}$	k_B	$K_B \cdot \frac{1}{[I]_0^4}$	$k_{Q_{L0}/Q_{S0}}$	$\frac{[Q_L]_0}{[Q_S]_0}$
$k_{Q_{AR},R}$	$N \cdot \frac{K_{Q_{AR},R}}{\mu \cdot [Q_{AR}]_0}$	λ_{x_1}	$\frac{\lambda_{Q_S}}{\mu}$	$k_{Q_{S0}/X_{AR0}}$	$\frac{[Q_S]_0}{[X_{AR}]_0}$

Based on mass-action kinetics the rate of formation of Q, Q_S and Q_L are given by Equations 5.17-5.19. The rate of generation of full length Q RNA is given by k_Q (Equation 5.13). We assume X_{AR} interacts with Q to form of Q_S RNA as shown by the second term in Equation 5.17, the remaining Q transcript continues to elongate at rate proportional to Q with a first order rate constant k_{Q_L} (Equation 5.17 and 5.19). Degradation and dilution due to growth is considered to follow first order kinetics as shown by the last term in Equations 5.18-5.20.

$$\frac{d[Q]}{dt} = k_Q - K_{AR}[X_{AR}][Q] - k_{Q_L}[Q] \qquad 5.17$$

$$\frac{d[Q_S]}{dt} = K_{AR}[X_{AR}][Q] - (\lambda_{Q_S} + \mu)[Q_S] \qquad 5.18$$

$$\frac{d[Q_L]}{dt} = k_{Q_L}[Q] - (\lambda_{Q_L} + \mu)[Q_L] \qquad 5.19$$

$$\frac{d[X_{AR}]}{dt} = k_{X_{AR}} - K_{AR}[X_{AR}][Q] - (\lambda_{X_{AR}} + \mu)[X_{AR}] \qquad 5.20$$

The nascent Q RNA, is an intermediate species in the cell (hence no dilution or degradation term is considered), which is either converted to Q_S or Q_L. We thus apply pseudo Steady State assumption to Equation 5.17 to yield an expression for Q as shown in Equation 5.21.

$$[Q] = \frac{k_Q}{k_{Q_L} + K_{AR}[X_{AR}]} \qquad (5.21)$$

We define parameter; $K_{Q-X_{AR}} = \frac{K_{AR}}{k_{Q_L}}$. Substituting Equations 5.21, 5.13 and 5.16 and into 5.18-5.20 yields the corresponding ODEs as shown in Equations 5.25-5.27.

We next consider the RNA interaction between transcripts X and Q_{AR}, Q_{AR} and X_{AR}. Based on mass action kinetics the rate of formation of X and Q_{AR} are given by Equations 5.21-5.22, where λ_X and $\lambda_{Q_{AR}}$ denote the degradation rate constant for X and Q_{AR} RNA respectively and μ denotes the specific growth rate. The formation of sense: antisense RNA complex between Q and X_{AR} RNA is assumed to follow irreversible second order kinetics with K_{AR} as 1×10^{-6} $M^{-1}s^{-1}$ (Shokeen *et al.*, 2010) as shown in Equation 5.21-5.22. We assume hybrid RNA complex of X and Q_{AR} renders X to become unavailable for translation. Since X_{AR} represents 5' ends of all P_X transcripts that possess stem loops I and II (truncated RNA between 104-223 nt and Anti-Q RNA), interaction of Q_{AR} with X_{AR} (hence also accounts for interaction of Q_{AR} with X) is considered in the second term of Equation 5.23.

$$\frac{d[X]}{dt} = k_X - K_{AR}[X]\cdot[Q_{AR}] - (\lambda_X + \mu)[X] \qquad (5.22)$$

$$\frac{d[Q_{AR}]}{dt} = k_{Q_{AR}} - K_{AR}[X_{AR}]\cdot[Q_{AR}] - (\lambda_{Q_{AR}} + \mu)[Q_{AR}] \qquad (5.23)$$

Substituting Equations 5.14 and 5.15 and into 5.22-5.23 yields the corresponding ODE's as shown in Equations 5.28-5.29.

5.5 A model for the regulation of conjugation initiation in the prgQ/prgX system

Combining the formulation from the previous sections, a set of mass action equation describing the balance of various transcripts, signaling molecules cCF10 and

iCF10 are listed (Equations 5.24-5.31). The dynamics of extracellular iCF10 (i), intracellular iCF10 (I) and cCF10 (C) is shown in Equations 5.30-5.32. The transport of signaling molecules cCF10 and iCF10 across the membrane protein PrgZ is modeled as a first order reaction (Equation 5.31-5.32) based on results reported in literature (Leonard *et al.*, 1996; Mehra *et al.*, 2008). It is assumed that the levels of PrgZ remain constant during the process of induction. Generation of extracellular iCF10 is modeled as first order with respect to both Q_S and Q_L but not from Q_{AR} as it assumed that truncated RNA does not participate in translation. The effect of endogenous cCF10 released by donor cells is considered negligible due to cCF10-sequestering activity of PrgY protein (Chandler *et al.*, 2005). Degradation and dilution due cell growth is considered for all intracellular species following first order kinetics denoted by the last term in Equations 5.25-5.29 and 5.31-5.32 (Mehra *et al.*, 2008). Dilution due to growth is not considered for extracellular iCF10 in Equation 5.30, however extracellular degradation is considered. The concentration of extracellular cCF10 is fixed for the steady state simulations.

$$[O] = \frac{N \cdot [I]^4}{[I]^4 + K_B \cdot [C]^4} \tag{5.24}$$

$$\frac{d[Q_S]}{dt} = K_{Q,R}[O] + K_{Q,D}([N]-[O]) \left(\frac{K_{Q-X_{AR}} \cdot [X_{AR}]}{1 + K_{Q-X_{AR}} \cdot [X_{AR}]} \right) - (\lambda_{Q_S} + \mu)[Q_s] \tag{5.25}$$

$$\frac{d[Q_L]}{dt} = K_{Q,R}[O] + K_{Q,D}([N]-[O]) \left(\frac{1}{1 + K_{Q-X_{AR}} \cdot [X_{AR}]} \right) - (\lambda_{Q_L} + \mu)\ Q_L \tag{5.26}$$

$$\begin{aligned} \frac{d[X_{AR}]}{dt} &= K_{X_{AR},R}[O] + K_{X_{AR},D}([N]-[O]) - K_{AR}[X_{AR}][Q_{AR}] \\ &\quad -[K_{Q,R}[O] + K_{Q,D}([N]-[O])] \left(\frac{K_{Q-X_{AR}} \cdot [X_{AR}]}{1 + K_{Q-X_{AR}} \cdot [X_{AR}]} \right) - (\lambda_{X_{AR}} + \mu)[X_{AR}] \end{aligned} \tag{5.27}$$

$$\frac{d[X]}{dt} = K_{X,R}[O] + K_{X,D}([N]-[O]) - K_{AR}[X]\cdot[Q_{AR}] - (\lambda_X + \mu)[X] \quad (5.28)$$

$$\frac{d[Q_{AR}]}{dt} = [K_{Q_{AR},R}[O] + K_{Q_{AR},D}([N]-[O])] - K_{AR}[X_{AR}]\cdot[Q_{AR}] - (\lambda_{Q_{AR}} + \mu)[Q_{AR}] \quad (5.29)$$

$$\frac{d[i]}{dt} = K_i([Q_s]+[Q_L])\cdot V_{conv} - K_{T_i}\cdot[i] - \lambda_i\cdot[i] \quad (5.30)$$

$$\frac{d[I]}{dt} = K_{T_i}[i] - (\lambda_I + \mu)[I] \quad (5.31)$$

$$\frac{d[C]}{dt} = K_{T_c}[c] - (\lambda_C + \mu)[C] \quad (5.32)$$

5.6 Dimensionless ODE model and model reduction for testing robustness of bistability

The set of equations describing the system behavior was converted to dimensionless equations by normalizing each variable to a reference OFF (repressed) state. The dimensionless time for reference state was chosen to be 1/μ. Dimensionless intracellular cCF10 (C/C_0 =1) and extracellular concentration (c/c_0=1) were set to one. The dimensionless form of parameters in Equations 5.24-5.32 is shown in Table 5.2. The dimensionless equations are shown in Equations 5.33-5.41.

$$\alpha = \frac{x_7^4}{x_7^4 + k_B x_8^4} \quad (5.33)$$

$$\frac{dx_1}{d\tau} = k_{Q,R}\ \alpha + k_{Q,D}(1-\alpha)\left(\frac{k_{Q-X_{AR}}x_3}{1+k_{Q-X_{AR}}x_3}\right) - (\lambda_{x_1}+1)x_1 \quad (5.34)$$

$$\frac{dx_2}{d\tau} = \frac{k_{Q,R}}{k_{Q_{L_0}/Q_{S_0}}}\ \alpha + k_{Q,D}(1-\alpha)\left(\frac{1}{1+k_{Q-X_{AR}}x_3}\right) - (\lambda_{x_2}+1)x_2 \qquad (5.35)$$

$$\frac{dx_3}{d\tau} = k_{X_{AR},R}\ \alpha + k_{X_{AR},D}(1-\alpha)\ - k_{AR}\cdot x_3\cdot x_5$$
$$-k_{Q_{S_0}/X_{AR_0}}k_{Q,R}\ \alpha + k_{Q,D}(1-\alpha)\left(\frac{k_{Q-X_{AR}}x_3}{1+k_{Q-X_{AR}}x_3}\right) - (\lambda_{x_3}+1)x_3 \qquad (5.36)$$

$$\frac{dx_4}{dt} = k_{X,R}\ \alpha + k_{X,D}(1-\alpha)\ - k_{AR}\cdot x_4\cdot x_5 - (\lambda_{x_4}+1)x_4 \qquad (5.37)$$

$$\frac{dx_5}{dt} = k_{Q_{AR},R}\ \alpha + k_{Q_{AR},D}(1-\alpha)\ - k_{AR}\cdot x_3\cdot x_5 - (\lambda_{x_5}+1)x_5 \qquad (5.38)$$

$$\frac{dx_6}{d\tau} = k_i k_{V_{conv}}(x_1 + k_{Q_{L_0}/Q_{S_0}}x_2) - k_{T_i}x_6 - \lambda_{x_6}x_6 \qquad (5.39)$$

$$\frac{dx_7}{d\tau} = k_{T_i}x_6 - (\lambda_{x_7}+1)x_7 \qquad (5.40)$$

$$\frac{dx_8}{d\tau} = k_{T_c} - (\lambda_{x_8}+1)x_8 \qquad (5.41)$$

To obtain steady state solution for the above set of dimensionless equations describing the system $\frac{dx_i}{dt}$ was set to 0, and the values of all species, x_i and α, were set to 1, corresponding to an OFF state, where i =1, 2, …8 are species summarized in Table 5.2. Using the initial OFF steady state assumption a set of algebraic equations for each variable was obtained as shown in Equations 5.42-5.47. The bistable behavior of the system was retained even when interaction between X_{AR} and Q_{AR} species was eliminated from the model. Hence, in the reduced model we eliminated the second term ($k_{AR}\cdot x_3\cdot x_5$) in Equation 5.36. Consequently, both X and Q_{AR} behave independently of

Q_L, Q_S and X_{AR} species and do not affect the bistable response of Q_L to cCF10. As a result the two corresponding dimensionless Equations 5.37 and 5.38 were not considered for further analysis.

$$from\ 5.34: k_{Q,R} = \frac{(1+k_{Q-X_{AR}})(\lambda_{x_1}+1)}{k_{Q-X_{AR}}} \tag{5.42}$$

$$from\ 5.35: \frac{k_{Q,R}}{k_{Q_{L_0}/Q_{S_0}}} = (\lambda_{x_2}+1)(1+k_{Q-X_{AR}}) \tag{5.43}$$

$$from\ 5.36: k_{Q_{S_0}/X_{AR_0}} k_{Q,R} = \frac{(k_{X_{AR},R}-\lambda_{x_3}-1)(1+k_{Q-X_{AR}})}{k_{Q-X_{AR}}} \tag{5.44}$$

$$from\ 5.39: k_i = \frac{(k_{T_i}+\lambda_{x_6})}{k_{V_{conv}}\ k_{Q_{L_0}/Q_{S_0}}+1} \tag{5.45}$$

$$from\ 5.40: k_{T_i} = (\lambda_{x_7}+1) \tag{5.46}$$

$$from\ 5.41: k_{T_c} = (\lambda_{x_8}+1) \tag{5.47}$$

Substituting equations 5.42-5.47 into dimensionless Equations 5.33-5.41 and excluding Equations 5.37-5.38 reduced the number of equations from 9 to 5 and yielded the final set of steady state equations 5.48-5.53 that were used to search the parameter space for bistability. This approach significantly reduced the number of parameters from about thirty in equations 5.33-5.41 to six in Equations 5.48-5.53, namely $k_{Q,D}$, $k_{X_{AR},R}$, $k_{X_{AR},D}$, $k_{Q_{L_0}/Q_{S_0}}$, $k_{Q-X_{AR}}$ and λ_{x_3}.

$$x_1 = [\alpha + k_{Q,D}(1-\alpha)]\left(\frac{1+k_{Q-X_{AR}}}{1+k_{Q-X_{AR}}x_3}\right)x_3 \qquad (5.48)$$

$$x_2 = [\alpha + k_{Q,D}(1-\alpha)]\left(\frac{1+k_{Q-X_{AR}}}{1+k_{Q-X_{AR}}x_3}\right) \qquad (5.49)$$

$$x_3 = \frac{(\wp^2 + 4\cdot k_{X_{AR},R}\cdot[\alpha + k_{X_{AR},D}(1-\alpha)]\cdot k_{Q-X_{AR}}\cdot(\lambda_{x_3}+1))^{1/2} - \wp}{2\cdot k_{Q-X_{AR}}\cdot(\lambda_{x_3}+1)} \qquad (5.50)$$

where $\wp = \{[\alpha + k_{Q,D}(1-\alpha)]\cdot(k_{X_{AR},R} - \lambda_{x_3} - 1)\cdot(k_{Q-X_{AR}}+1)$
$-k_{X_{AR},R}\cdot[\alpha + k_{X_{AR},D}(1-\alpha)]\cdot k_{Q-X_{AR}} + (\lambda_{x_3}+1)\}$

$$x_6 = \frac{x_1}{(k_{Q_{L_0}/Q_{S_0}}+1)} + \frac{k_{Q_{L_0}/Q_{S_0}}\cdot x_2}{(k_{Q_{L_0}/Q_{S_0}}+1)} \qquad (5.51)$$

$$x_7 = x_6 \qquad (5.52)$$

$$x_8 = 1 \qquad (5.53)$$

5.7 Robustness of bistability

The parameter values used to obtain the steady state behavior shown in Figure 4.5 C-D is summarized in Table 5.1 and 5.4. As described above, the final set of dimensionless parameters reduced to six, namely $k_{Q,D}$, $k_{X_{AR},R}$, $k_{X_{AR},D}$, $k_{Q_{L_0}/Q_{S_0}}$, $k_{Q-X_{AR}}$ and λ_{x_3} (described in Table 5.2) in the Equations 5.48-5.53. Among the six parameters, the value of $k_{Q,D}$, $k_{X_{AR},D}$ and λ_{x_3} used in the model simulation were obtained from experimental data. Parameter $k_{Q,D}$ is given by, $k_{Q,D} = \frac{K_{Q,D}}{K_{Q,R}}$, which is the ratio of

transcription rate of Q RNA under derepressed to repressed state. The value of $k_{Q,D} = 9$ was obtained from *lacZ* reporter study in Figure 4.2, The transcription rate of X_{AR} under derepressed and repressed state obtained from the RNAP collision model, was used to calculate, $k_{X_{AR},D} = \frac{K_{X_{AR},D}}{K_{X_{AR},R}}$. The transcription rates summarized in Table 5.3, were used to estimate $k_{X_{AR},D}$=0.12. The value of $\lambda_{x_3} = \frac{\lambda_{X_{AR}}}{\mu}$, which is the ratio of X_{AR} to the specific growth rate, was determined next. The half-life of Anti-Q has been reported to be approximately 85 min (Bae *et al.*, 2002), a similar half-life for X_{AR} was used in our model, thus setting the degradation rate $\lambda_{X_{AR}}$ to 1.36×10^{-4} s^{-1}. The specific growth rate was experimentally determined to be 2.56×10^{-4} s^{-1} $\left(i.e.\ \mu = \frac{\ln 2}{(45\,\text{min})} \right)$. The resulting ratio $\lambda_{x_3} = \frac{\lambda_{X_{AR}}}{\quad}$, was set at 0.53.

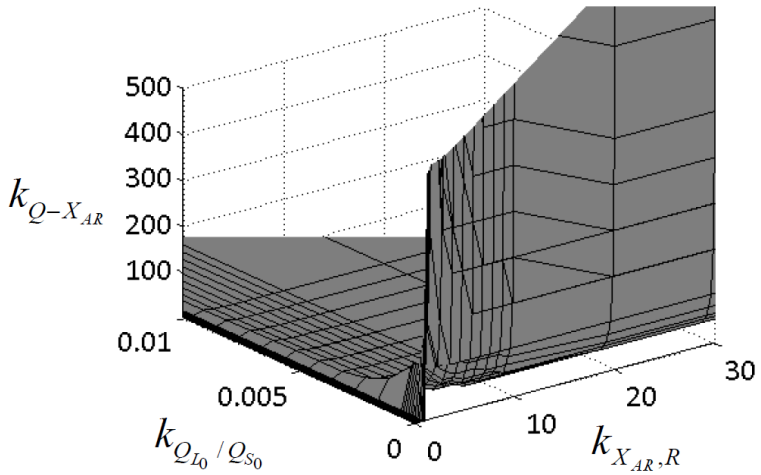


Figure 5.2: Parameters space for dimensionless parameters $k_{Q_{L_0}/Q_{S_0}}$, $k_{X_{AR},R}$ and $k_{Q-X_{AR}}$ where bistability is predicted for RC/AR case.

Table 5.3 The parameter values used for RC/AR, RC⁻/AR⁻, RC⁻/Anti-Q and RC/AR⁻ cases (values used for Fig 4.5 D)

PARAMETERS	RC/AR	RC/AR⁻	RC⁻/AR⁻	RC⁻/Anti-Q	Units
$K_{Q,R}$	7.23×10^{-3}	7.23×10^{-3}	1.10×10^{-2}	1.10×10^{-2}	s^{-1}
$K_{Q,D}$	8.87×10^{-2}	8.87×10^{-2}	1.00×10^{-1}	1.00×10^{-1}	s^{-1}
$K_{X_{AR},R}$	1.02×10^{-2}	1.02×10^{-2}	0	0	s^{-1}
$K_{X_{AR},D}$	1.21×10^{-3}	1.21×10^{-3}	0	0	s^{-1}
$K_{X,R}$	8.23×10^{-3}	8.23×10^{-3}	1.20×10^{-2}	1.20×10^{-2}	s^{-1}
$K_{X,D}$	1.21×10^{-3}	1.21×10^{-3}	1.20×10^{-2}	1.20×10^{-2}	s^{-1}
$K_{Q_{AR},R}$	1.80×10^{-3}	1.80×10^{-3}	0	0	s^{-1}
$K_{Q_{AR},D}$	1.08×10^{-2}	1.08×10^{-2}	0	0	s^{-1}
$K_{Q-X_{AR}}$	1×10^{4}	0	0	--	$nM^{-1}\ s^{-1}$
K_{AR}	1×10^{-3}	0	0	0	$nM^{-1}\ s^{-1}$
$K_{Q-Anti-Q}$ (Anti-Q only case)	--	--	--	1×10^{4}	$nM^{-1}\ s^{-1}$
$K_{Anti-Q,R}$ (Anti-Q only case)	--	--	--	1.20×10^{-2}	s^{-1}
$K_{Anti-Q,D}$ (Anti-Q only case)	--	--	--	1.20×10^{-2}	s^{-1}

Table 5.4 Parameter range scanned for finding bistability for wildtype RC/AR case, RC$^-$/AR$^-$, RC/AR$^-$ and RC$^-$/Anti-Q (values used for Fig 4.5 D)

Range of dimensionless parameters for which bistability is observed for wildtype RC/AR case		
Dimensionless parameter	**Nominal values for which bistability is observed**	**Range of parameter space scanned with respect to Nominal set for which bistability is observed**
$k_{Q_{L_0}/Q_{S_0}}$	7.448	1×10^{-4} -3×10^{-2}
$k_{X_{AR},R}$	1.44×10^{-4}	1-100
$k_{Q-X_{AR}}$	86.9	1×10^{-3} -1×10^{3}
Range of dimensionless parameters scanned for RC$^-$/AR$^-$, RC/AR$^-$ and RC$^-$/Anti-Q to test presence of bistability. No bistability was observed in this range.		
Dimensionless parameter	**Nominal set (no bistability observed)**	**Range of parameter space scanned with respect to Nominal set (no bistability observed)**
$k_{Q_{L_0}/Q_{S_0}}$	7.448	1×10^{-3} -1×10^{4}
$k_{X_{AR},R}$	1.44×10^{-4}	1×10^{-3} -1×10^{4}
$k_{Q-X_{AR}}$	86.9	1×10^{-3} -1×10^{4}

To assess the robustness of the bistable behavior the parameter space for the remaining three parameters $k_{X_{AR},R}$, $k_{Q_{L_0}/Q_{S_0}}$, and $k_{Q-X_{AR}}$ was simultaneously searched. The parameter space for these parameters was coarsely grained and searched simultaneously to check for multiple steady states using MATLAB function polyxpoly. For the data shown in Figure 4.5 C-D, the corresponding parameter set used to obtained bistability were, $k_{X_{AR},R}$ =7.448, $k_{Q_{L_0}/Q_{S_0}}$ =1.44×10^{-4} and $k_{Q-X_{AR}}$ =86.9. Bistability was observed for a wide range of parameter values spanning over two orders in magnitude (Table 5.4, Figure

5.2), suggesting the bistable behavior observed is robust. We denote this parameter set as X_0.

5.8 De-coupling RNAP collision and antisense interaction in the model

To evaluate bistable behaviour under RC and RC$^-$, and AR and AR$^-$ conditions, equations and/or parameter values were altered as shown in Table 5.4. For AR$^-$, equations corresponding to X_{AR} and Q_{AR} were removed from the model and the second order RNA interaction rate constant K_{AR} and Equilibrium constant of the interaction between sense and antisense RNA, $K_{Q-X_{AR}}$ were set to zero. For the RC$^-$ case, the transcription rates of truncated RNA were set to zero, and that of the full length RNA was set to intrinsic RNAP firing rate at the corresponding promoters. For the cases RC/AR$^-$ (only RNAP collision, but no antisense interaction) and RC$^-$/AR$^-$ (neither effects are considered) the rate constants used in the model are summarized in Table 5.3.

The effect of Anti-Q alone without RC on bistability was evaluated. In this case (RC$^-$ /Anti-Q), antisense interaction from only Anti-Q is considered. Since truncated RNA is not generated in this case, $K_{X_{AR},R}$ $K_{X_{AR},D}$, $K_{Q_{AR},R}$ and $K_{Q_{AR},D}$ was set to zero. We assume that full length of X RNA is processed to give Anti-Q (Bae *et al.*, 2002), such that $K_{Anti-Q,R} = K_{Anti-Q,D} = K_{X,R} = K_{X,D} = 1.2\times10^{-2}\ s^{-1}$. The equilibrium constant for Anti-Q interaction with nascent Q RNA was the same as used for X_{AR}, i.e.

$K_{Q-Anti-Q} = K_{Q-X_{AR}}$ =1×10^{-3} (nM.s)$^{-1}$ (Table 5.3). Despite restoring antisense effect of Anti-Q, bistability was not observed.

For the cases of RC$^-$/AR$^-$, RC/AR$^-$ and RC$^-$/Anti-Q, the parameter space was searched rigorously for X_0×10^{-3} to X_0×10^4, where X_0 is the parameter set for which bistable behaviour is shown in wildtype RC/AR case as summarized in Table 5.4. Despite the wide range of parameter space was searched, bistability was not observed in any of these cases.

5.9 Concluding remarks

The model described in this chapter is used and extended in subsequent chapters for analysis of pCF10 system under different experimental conditions.

CHAPTER 6

ROLE OF pCF10 PLASMID COPY NUMBER IN THE PRGQ-PRGX GENETIC SWITCH IN BIOFILM CELLS

6.1 Summary

Conjugation is an important mode of horizontal gene transfer in bacteria, enhancing the spread of antibiotic resistance. Here we demonstrate that higher plasmid copy number of pCF10 plasmid in biofilm state increases robustness of a genetic switch controlling conjugative gene transfer in *E. faecalis*. Mathematical modeling suggested that a higher plasmid copy number in biofilm cells would enhance a bistable switch in conjugation regulation featuring a delayed, but increased response to the mating signal, both of which were observed experimentally. Experimental data indicates a bimodal response to induction of conjugation in populations of plasmid-containing donor cells in biofilms, however, less pronounced in planktonic bacterial cultures. Both mathematical model and experimental data point towards population heterogeneity occurs due to variation in pCF10 plasmid copy among donor population. Such heterogeneity in plasmid copy number can have important implications on the response of pCF10 cells to induction and hence conjugation.

6.2 Introduction

Biofilm is defined as an aggregate of bacterial cells adhering to each other on a surface, often embedded within an extracellular matrix of DNA, proteins and polysaccharides (Costerton *et al.*, 1999; Watnick and Kolter, 2000). Microbial cells growing in biofilms have been shown to behave differently than planktonic counterparts, i.e. single-cells growing in suspension. In microbial biofilms, cell-cell communication and coordinated multicellular interactions are of paramount importance. Quorum sensing and related forms of intercellular signaling have been shown to play critical roles in biofilm development and dispersal (Boles and Horswill, 2008). In addition to communication by extracellular signals, the biofilm environment may contribute to microbial evolution by serving as an important niche for horizontal genetic transfer by transformation or conjugation (Nguyen *et al.*, 2010) as has been previously showed that an *in vivo* biofilm, namely an endocarditis vegetation, could serve as niche for high-frequency transfer of pCF10 (Hirt *et al.*, 2005).

In this chapter, we investigate the role of pCF10 plasmid copy number between biofilm and planktonic cells. A *gfp* reporter was transcriptionally fused in-frame to *prgB*, the gene encoding Aggregation Substance (Asc10) in the pCF10 conjugative plasmid was used to study the pheromone-inducible response of donor cells (Cook *et al.*, 2011) (Figure 6.1). As Asc10 production facilitates conjugative transfer of pCF10, the expression of GFP is used as a reporter of conjugation readiness. Using this reporter strain, a significant difference in the population dynamics of the pheromone response

between biofilm and planktonic growth conditions was observed, consistent with the hypothesis that biofilm growth results in formation of distinct cell types that impact the behavior and regulation of pCF10 transfer.

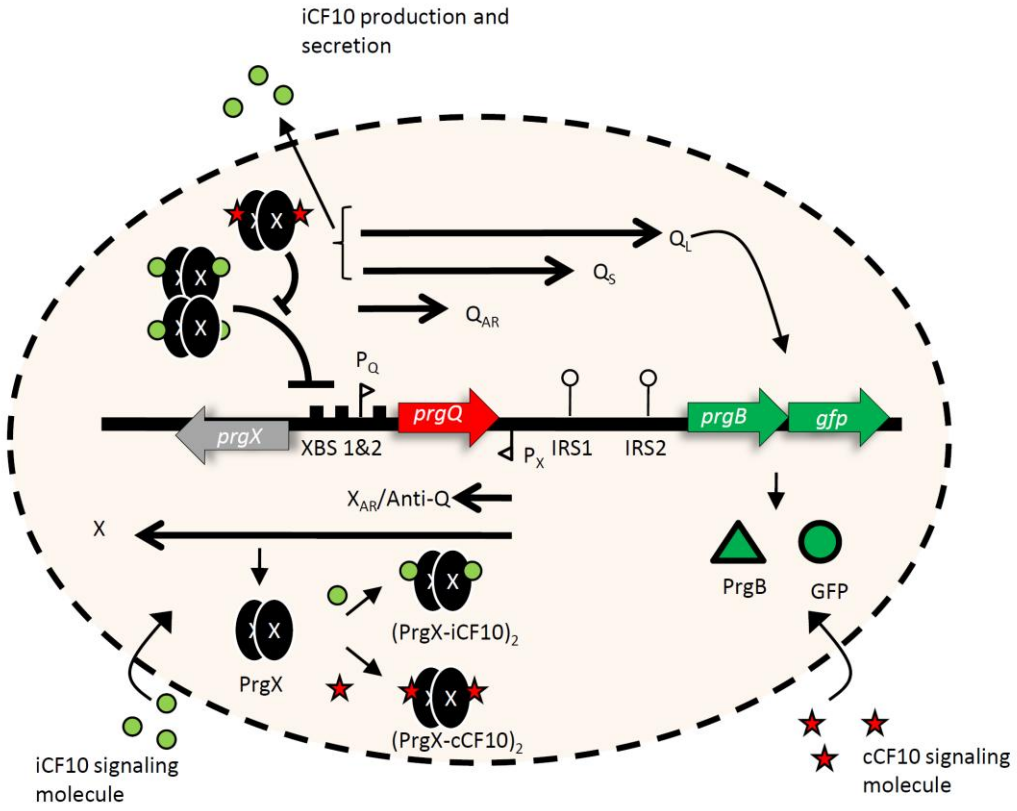


Figure 6.1: The pCF10 gene-regulatory network considered for the mathematical model. PrgB protein expressed as a result of induction includes expression of *gfp* gene fused in-frame to *prgB.*

6.3 Results

6.3.1 Mathematical modeling of effect of plasmid pCF10 copy number on response to induction

To explore possible causes of the differing response to conjugative peptide induction, we employed the mathematical model for the pCF10 genetic network discussed in Chapter 4 and 5 (Figure 6.1). The *prgQ-prgX* gene pair controls genetic switch for the onset of conjugation in the pCF10 based system. Recent work has demonstrated that transcriptional interference and antisense RNA regulation due to overlapping configuration of *prgQ* and *prgX* operons, confers the pCF10 system a bistable switch like behavior in response to induction with pheromone (Chatterjee *et al.*, 2011). Transcriptional interference (TI) occurs due to collision between elongating RNAP within the overlapping DNA between promoters P_Q and P_X and causes generation of shorter truncated RNA within the overlapping locus. This results in mutual suppression of transcription from both promoters, with expression from the aggressive promoter prevailing over the weak one (Shearwin *et al.*, 2005). In addition to TI, the sense and antisense truncated and full-length RNA originating from this overlapping locus have been shown to repress each other respectively (Chatterjee *et al.*, 2011) due to interaction between various stem loop structures located on P_Q transcripts and P_X transcripts (Shokeen *et al.*, 2010). The resultant regulatory circuit for this system is summarized in Figure 6.2 A. The mutual inhibition via both TI and antisense regulation

results in a Double-negative feedback loop (Figure 6.2 A) contributing to the bistable response of this system.

Based on the gene regulatory network shown in Figure 6.1, a set of nine ordinary differential equations (ODEs) based on mass-action kinetics for RNA species Q_S, Q_L, Q_{AR}, X and X_{AR}, PrgB protein and iCF10 and cCF10 described in Equations 6.1-6.9 (Table 6.1). The notation of all species and the parameters used for the simulation are summarized in Table 6.2. The ODEs describing transcripts Q_S, Q_L, Q_{AR}, X and X_{AR} in Equations 6.1-6.5 are a balance of transcription rate, rate of removal due to antisense interactions, degradation and dilution due to growth. Protein PrgB shown in Equation 6.6, is balance of rate of PrgB production, degradation and dilution terms. The dynamics of extracellular iCF10 (i), intracellular iCF10 (I) and cCF10 (C) is shown in Equations 6.7-6.9. The transport of signaling molecules cCF10 and iCF10 across the cell membrane is modeled as a first order reaction (Equation 6.8-6.9). Generation of extracellular iCF10 is modeled as first order with respect to both Q_S and Q_L RNA. Dilution due to growth is not considered for extracellular iCF10, however, extracellular degradation is considered in Equation 6.7.

The ODEs shown in Equation 6.1-6.9 were solved for steady state and dynamic solution for fixed extracellular concentration of cCF10 (Figure 6.2 B-E). A characteristic S-shaped bistable response of PrgB to extracellular cCF10 is shown in Figure 6.2 B. The bistable curve is comprised of three sections: the lower and upper part of the curve corresponds to “OFF” (low level of PrgB) and “ON” states (high level of PrgB) of the

donor cell. The middle section is characterized by multiple steady states, two stable states corresponding to "ON" and "OFF" and an unstable steady state which is not observed experimentally. The separation between "OFF" and "ON" state is characterized by the width of the bistable region, i.e. the region with multiple steady states (Figure 6.2 B). However, such a bistable behavior is intrinsic to the gene regulatory network and would be present both in planktonic cells as well as biofilm cells.

Mathematical modeling suggested that an increase in the copy number of pCF10 could greatly alter induction responses to cCF10 (Figure 6.2 C-H). The effect of increased plasmid copy number manifests itself through increased number of operator sites (XBS 1 and 2) available for iCF10 bound PrgX tetramers and cCF10 bound PrgX dimers to bind (Equation 10) and an increase in the number of copies of Q_S and Q_L transcripts per donor cell, which consequently results in increased production of inhibitor iCF10. Donor cells with fewer copies of pCF10 are predicted to require lower amounts of cCF10 to turn "ON" and also respond at a lower PrgB expression level than cells with a higher copy number (Figure 6.2 C). Interestingly, increasing plasmid copy number widens the bistable region, suggesting that the "ON" and "OFF" populations are better separated.

Similarly, the model predicts that cells with high plasmid copy number respond slowest to induction and at a higher concentration of cCF10 whereas low copy number cells respond faster and at a lower concentration of cCF10 (Figure 6.2 D-E). A broader plasmid copy number distribution is predicted to give rise to a bimodal population

response at longer exposure time to low levels of inducer, such that cells with high plasmid copy number continue to exist in the "OFF" state even after long exposure time to inducer, whereas cells with lower copy number switch to "ON" state (Figure 6.2 D-E and Figure 6.2 G-H), consistent with experimental data shown in Figure 6.3 (Cook *et al.*, 2011). Our model predicts that a higher pCF10 copy number and copy number heterogeneity would enhance a bimodal response to induction with pheromone.

Table 6.1: List of ODEs used in the mathematical model

6.1	$\frac{d[Q_S]}{dt} = K_{Q,R}[O] + K_{Q,D}([N]-[O])\left(\frac{K_{Q-X_{AR}}\cdot[X_{AR}]}{1+K_{Q-X_{AR}}\cdot[X_{AR}]}\right) - (\lambda_{Q_S}+\mu)[Q_s]$
6.2	$\frac{d[Q_L]}{dt} = K_{Q,R}[O] + K_{Q,D}([N]-[O])\left(\frac{1}{1+K_{Q-X_{AR}}\cdot[X_{AR}]}\right) - (\lambda_{Q_L}+\mu)\ Q_L$
6.3	$\frac{d[Q_{AR}]}{dt} = [K_{Q_{AR},R}[O] + K_{Q_{AR},D}([N]-[O])] - K_{AR}[X_{AR}]\cdot[Q_{AR}] - (\lambda_{Q_{AR}}+\mu)[Q_{AR}]$
6.4	$\frac{d[X_{AR}]}{dt} = K_{X_{AR},R}[O] + K_{X_{AR},D}([N]-[O]) - K_{AR}[X_{AR}][Q_{AR}] - [K_{Q,R}[O] + K_{Q,D}([N]-[O])]\left(\frac{K_{Q-X_{AR}}\cdot[X_{AR}]}{1+K_{Q-X_{AR}}\cdot[X_{AR}]}\right) - (\lambda_{X_{AR}}+\mu)[X_{AR}]$
6.5	$\frac{d[X]}{dt} = K_{X,R}[O] + K_{X,D}([N]-[O]) - K_{AR}[X]\cdot[Q_{AR}] - (\lambda_X+\mu)[X]$
6.6	$\frac{d[\mathrm{Pr}\,gB]}{dt} = K_{PrgB}[Q_L] - (\lambda_{PrgB}+\mu)[PrgB]$
6.7	$\frac{d[i]}{dt} = K_i([Q_s]+[Q_L])\cdot V_{iCF10} - K_{T_i}\cdot[i]/V_{iCF10} - \lambda_i\cdot[i]$
6.8	$\frac{d[I]}{dt} = K_{T_i}[i].V_{iCF10} - (\lambda_I+\mu)[I]$
6.9	$\frac{d[C]}{dt} = K_{T_c}[c].V_{cCF10} - (\lambda_C+\mu)[C]$
6.10	$[O] = \frac{N\cdot[I]^4}{[I]^4 + K_B\cdot[C]^4}$

Table 6.2: List of parameters used for the mathematical model

Parameter	Description	Value used for bistability	Units	References
μ	Specific growth rate of donor cells	2.58×10^{-4}	s^{-1}	(Chatterjee *et al.*, 2011)
$K_{Q,U}$	Transcription rate constant of Q in uninduced state	7.50×10^{-3}	s^{-1}	
$K_{Q,I}$	Transcription rate constant of Q in induced state	8.87×10^{-2}	s^{-1}	
$K_{X_{AR},U}$	Transcription rate constant of X_{AR} in uninduced state	1.02×10^{-2}	s^{-1}	
$K_{X_{AR},I}$	Transcription rate constant of X_{AR} in induced state	1.20×10^{-3}	s^{-1}	
$K_{X,U}$	Transcription rate constant of X in uninduced state	1.2×10^{-3}	s^{-1}	
$K_{X,I}$	Transcription rate constant of X in induced state	1.20×10^{-3}	s^{-1}	
$K_{Q_{AR},U}$	Transcription rate constant of Q_{AR} in uninduced state	1.80×10^{-3}	s^{-1}	
$K_{Q_{AR},I}$	Transcription rate constant of Q_{AR} in induced state	1.53×10^{-3}	s^{-1}	
K_{PrgB}	Generation rate of PrgB protein	1×10^{-2}	s^{-1}	(Mehra *et al.*, 2008; Sotiropoulos and Kaznessis, 2007)
K_i	Generation rate of extracellular iCF10	1×10^{-3}	s^{-1}	(Chatterjee *et al.*, 2011)
V_{conv}	Volume conversion factor	1	-	
K_{T_i}	Transport rate constant of iCF10	1×10^{-3}	s^{-1}	
K_{T_C}	Transport rate constant of cCF10	1×10^{-3}	s^{-1}	
K_{AR}	Rate constant of interaction between X RNA and Q_{AR} RNA	1×10^{-3}	$(nM)^{-1}.s$	(Eguchi *et al.*, 1991; Shokeen *et al.*, 2010)
$K_{P_Q - X_{AR}}$	Equilibrium constant of Q and X_{AR} interaction	4.43	$(nM)^{-1}$	(Eguchi *et al.*, 1991; Shokeen *et al.*, 2010)
K_B	Equilibrium constant of DNA binding reaction	1×10^{8}	-	(Chatterjee, A., 2011, Nakayama, 1994)
λ_{Q_S}	Degradation rate of Q_S mRNA	1×10^{-3}	s^{-1}	(Sotiropoulos and Kaznessis, 2007)
λ_{Q_L}	Degradation rate of Q_L mRNA	1×10^{-1}	s^{-1}	(Sotiropoulos and Kaznessis, 2007)
$\lambda_{X_{AR}}$	Degradation rate of X_{AR} RNA	1.36×10^{-4}	s^{-1}	(Bae *et al.*, 2002)
$\lambda_{Q_{AR}}$	Degradation rate of Q_{AR} RNA	1×10^{-3}	s^{-1}	(Sotiropoulos and Kaznessis, 2007)
λ_X	Degradation rate of X RNA	1×10^{-4}	s^{-1}	(Sotiropoulos and Kaznessis, 2007)
λ_{PrgB}	Degradation rate of PrgB protein	1×10^{-6}	s^{-1}	(Mehra *et al.*, 2008)
λ_i	Degradation rate of extracellular iCF10	1×10^{-6}	s^{-1}	(Mehra *et al.*, 2008)
λ_I	Degradation rate of intracellular iCF10	1×10^{-6}	s^{-1}	(Mehra *et al.*, 2008)
λ_C	Degradation rate of intracellular cCF10	1×10^{-6}	s^{-1}	(Mehra *et al.*, 2008)

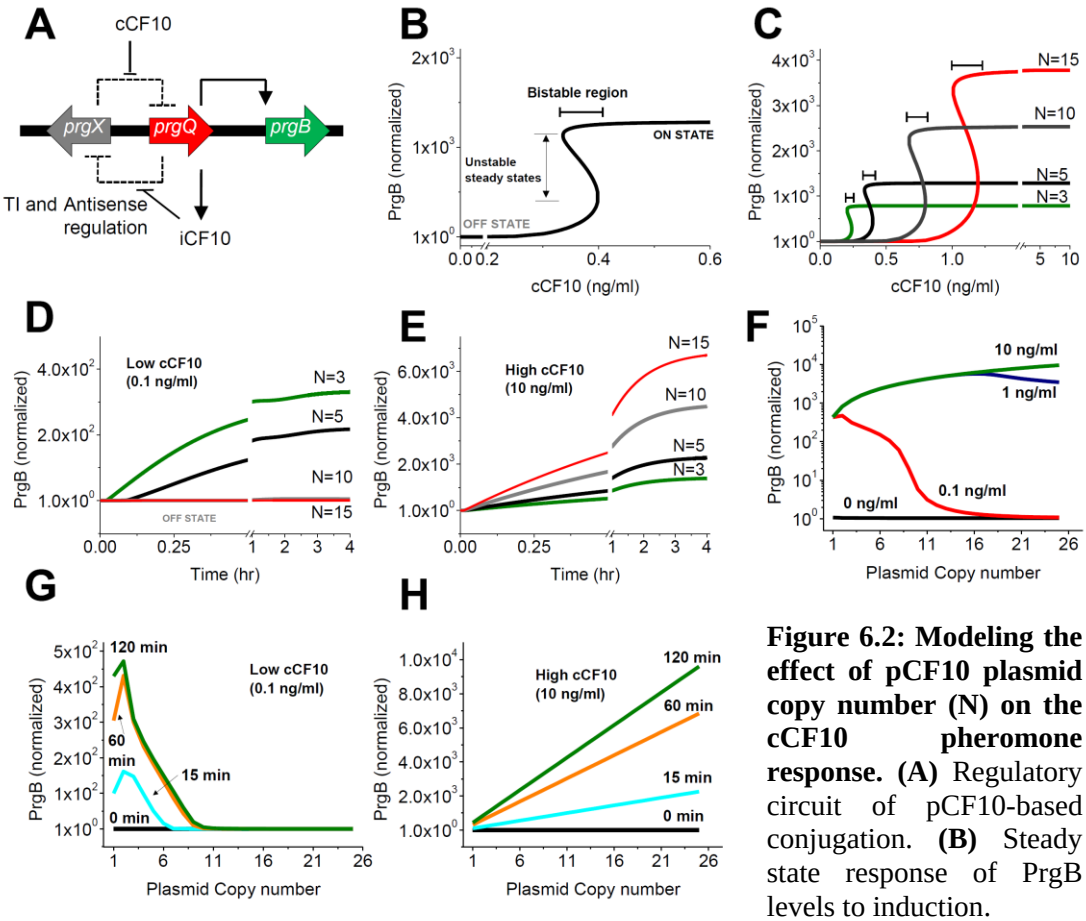


Figure 6.2: Modeling the effect of pCF10 plasmid copy number (N) on the cCF10 pheromone response. (A) Regulatory circuit of pCF10-based conjugation. **(B)** Steady state response of PrgB levels to induction. **(C)** Bistable switch response of pCF10 to induction with cCF10 for different plasmid copy numbers. The bistable region increases with plasmid copy number along with an increased threshold cCF10 concentration required to turn "ON" the genetic switch. PrgB levels shown in b-c are normalized to initial "OFF" steady state at 0 ng/mL of cCF10. **(D-E)** Dynamic response of donor cells with different plasmid copy numbers to low (0.1 ng/mL, D) and high (10ng/mL, E) induction concentrations of cCF10. Cells with high plasmid copy number respond slowest to induction and at a higher concentration of cCF10 whereas low copy number cells respond faster and at a lower concentration of cCF10. PrgB levels shown in D-E are normalized to initial state at time t=0 corresponding to steady state levels at 0 ng/mL of cCF10. **(F-H)** Steady state and dynamic response of donor cells with different pCF10 copy numbers. **(F)** Steady state response of PrgB levels to different levels of cCF10 induction (0-10 ng/ml) for different pCF10 copy numbers. PrgB levels are normalized to steady state levels corresponding to 0 ng/mL of cCF10. (B-C) Dynamic response of PrgB levels (0 to 120 min) to 1 ng/mL cCF10 induction at time t=0, for different pCF10 copy numbers. PrgB levels are normalized to initial condition corresponding to steady state levels at 0 ng/mL of cCF10.

6.3.2 Experimental data indicates different pattern of induction of pCF10 carrying donor cells between planktonic and biofilm cells

The response of pCF10 carrying donor cells to induction grown under planktonic and biofilm conditions was quantified by measuring expression of a GFP reporter which was transcriptionally fused in-frame to *prgB*, in the pCF10 conjugative plasmid (Figure 6.1 and Figure 6.3 A-D) . The experiments shown in Figure 6.3 were performed by Laura C. Cook in Prof. Gary M. Dunny's laboratory. The steady state response of the *prgQ-prgX* switch to pheromone titrations ranging from 0.1-10 ng/mL of cCF10 (for an induction time of 60 minutes) was compared between planktonic and biofilm cells as demonstrated by the Flow cytometric analysis shown in Figure 6.3 A-B. While planktonic cells demonstrated a homogenous unimodal distribution pattern in response to induction to different concentrations of inducer, biofilm cells showed a distinct bimodal response to induction.

Similarly, a time dependent unimodal increase in GFP expression was observed when identical levels of cCF10 (0.1 ng/mL) were used to induce planktonic cells for time periods ranging from 0-120 minutes (Figure 6.3 C-D). When cells growing in the biofilm state were similarly induced, a markedly different bi-modal induction pattern was observed for longer exposure times (Figure 6.3 D). Biofim cells demonstrated a slower response to induction compared to planktonic cells even though the growth rates of planktonic and biofilm cells were found to be similar (Figure 6.3 E).

The difference in response between the planktonic and biofilm cells in Figure 6.3 A-D, is consistent with the model prediction of higher pCF10 plasmid copy number in biofilm cells compared to planktonic cells. However, there could be additional factors which could give rise to bimodal response of biofilm cells, such as role played by the biofilm matrix in non-homogenous diffusion of cCF10 signal. However, this possibility was eliminated when a similar response to induction was observed for dispersed biofilm cells (data not shown). Additionally, loss of bimodal induction profile of a GFP fusion construct derived from pCF10 where transcription was driven by a single P_Q promoter construct, where the gene encoding pheromone receptor/conjugation repressor protein, PrgX, was deleted, indicated that the bimodal response was intrinsic to *prgQ-prgX* genetic switch (Figure 6.3 F). Adding *prgX in trans* rescued the bimodal response (Figure 6.3 G). This indicated that the bimodal distribution in GFP expression observed with the pheromone-inducible construct was a result of biofilm effects on the native *prgQ-prgX* genetic switch and not due to biofilm matrix structure.

6.3.3 Model predicts biofilm cells have a higher pCF10 copy number than planktonic cells

The steady state PrgB levels in response to different concentrations of cCF10 (low levels of cCF10: 0 and 0.1 ng/mL, medium level of cCF10: 1 ng/mL and high level of cCF10: 10 ng/mL) for different pCF10 copy numbers is shown in Figure 6.2 F-H. We consider a 10-fold increase in PrgB expression with respect to the initial off steady state (corresponding to 0 ng/mL of cCF10) as system switching to "ON" state. At low

concentrations of cCF10, cells are predicted to exist in the "OFF" state irrespective of the plasmid copy number (0 and 0.1ng/mL). At medium inducer levels (1 ng/mL), cells with low plasmid copy numbers are predicted to turn "on" PrgB expression levels. At high levels of induction (10 ng/mL), the model predicts that low and medium copy cells are turned "ON", however, cells with very high plasmid copy numbers continue to stay "OFF", until an even higher dose of inducer is administered. The results shown in Figure 6.2 C-F are consistent with experimental data shown in Figure 6.3 A-G.

We predict that biofilm cells demonstrate higher average plasmid copy number in addition to broader distribution of plasmid copy number compared to planktonic cells, which is consistent with experimental observations as shown in Figure 6.3 A-G. A subpopulation with higher plasmid copy number continues to be in the "OFF" state even at high levels of cCF10, whereas cells with lower plasmid pCF10 copy number switch to the "ON" state.

The temporal response of PrgB expression to low levels of induction (0.1 ng/mL) applied at time t=0 for different pCF10 copy numbers is shown Figure 6.2 D and Figure 6.2 G. Our model predicts that cells with low pCF10 copy numbers respond earliest to induction, whereas, cells with high copy number respond most slowly. The model prediction of a broader plasmid copy number distribution in biofilm cells entails a bimodal population response at longer exposure time to inducer, with high plasmid copy number cells continuing to be in the "OFF" state even after long exposure to inducer,

whereas cells with lower copy number switching to "ON" state, consistent with experimental data shown in Figure 6.3 C-D.

6.3.4 pCF10 copy number and heterogeneity is increased in biofilm cells

qPCR on genomic DNA obtained from uninduced planktonic and biofilm cells showed that the average pCF10 copy number of uninduced biofilm cells demonstrate a statistically significant increase in pCF10 copy number of 1.5-2 times than that of planktonic cells (Figure

6.3 H), which was also confirmed by pulse-field gel electrophoresis (Figure 6.3 I) (Cook *et al.*, 2011), consistent with the model prediction. Additionally, the mathematical model predicted that cells which continue to stay in the OFF state even after being exposed to inducer. qPCR on "ON" and "OFF" subpopulations of biofilm and planktonic cells in response to induction with 0.1 ng/mL of cCF10 (sorted based on GFP expression), showed that "OFF" cells (cells not expressing GFP) had a statistically higher copy number of pCF10 than "ON" cells (cells expressing GFP) during induction (Figure 6.3 J). On the other hand, planktonic cells induced with the same level of cCF10 did not shown any statistical difference between "ON" and "OFF" sub-populations (Figure 6.3 J). Additionally, while the induced biofilm cells showed similar copy number compared to planktonic cells (both induced and uninduced, approximately 3-5 copies per chromosome), uninduced biofilms showed a statistically much higher pCF10 copy number (8-15 copies per chromosome).

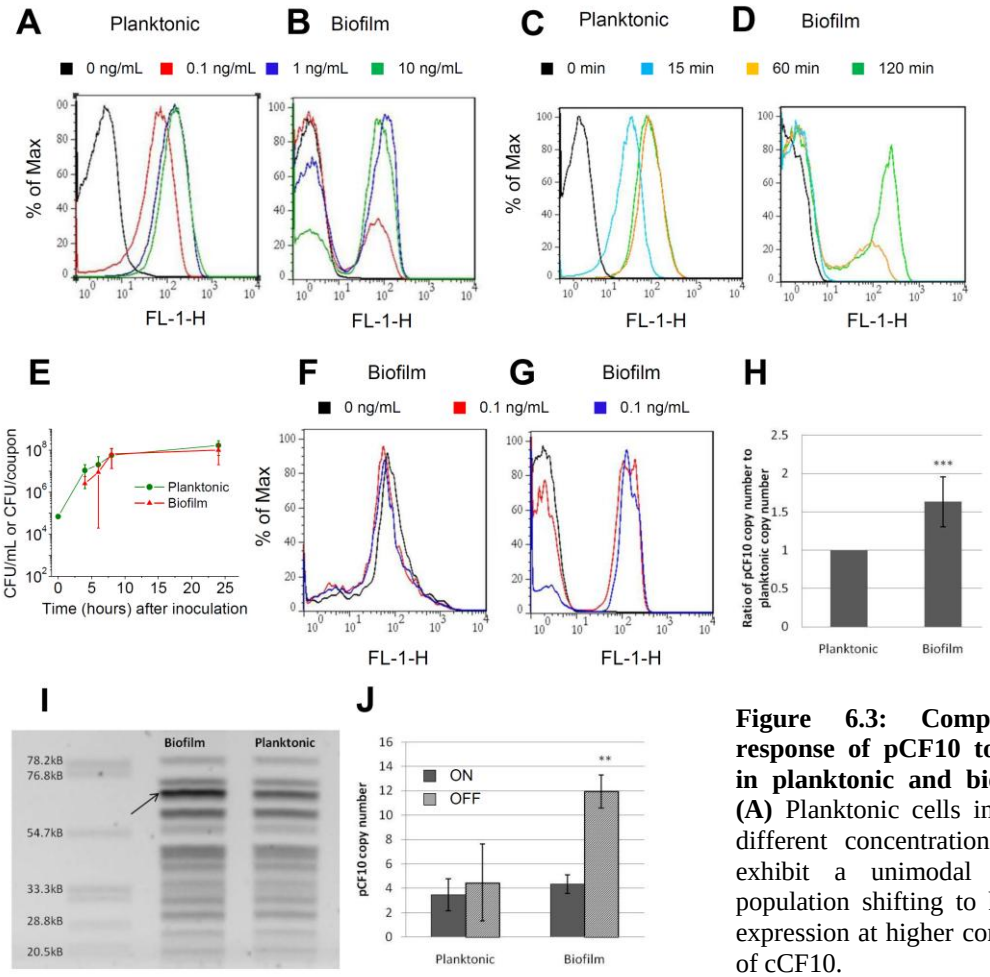


Figure 6.3: Comparison of response of pCF10 to induction in planktonic and biofilm cells. (A) Planktonic cells induced with different concentration of cCF10 exhibit a unimodal pattern of population shifting to higher GFP expression at higher concentrations of cCF10. **(C)** Cells grown in a liquid culture exhibit a unimodal shift pattern in induction (0.1 ng/mL of cCF10) with time. **(D)** A time course of biofilm cell induction (0.1 ng/mL of cCF10) shows bimodal induction kinetics. Horizontal axis in (A-D) = FL1 (GFP) expression, Vertical axis in (A-D) = % of maximum cell number. **(E)** Growth of biofilm and planktonic cells in the CDC Biofilm Reactor. (**F)** A constitutively expressed GFP reporter driven by single promoter P_Q (due to lack of *prgX* repressor gene) shows a unimodal "ON" state. **(G)** The same constitutive reporter plasmid demonstrates bimodal induction when the *prgX* is expressed *in trans* from the chromosome. Horizontal and vertical axes are the same as in (A-D). Error bars show standard deviation using n≥3 separate experiments. **(H)** Average pCF10 copy number in uninduced biofilm and planktonic cells by qPCR. **(I)** Pulse field gel electrophoresis (PFGE) of biofilm and planktonic cells shows that biofilm cells have a higher pCF10 DNA/chromosomal DNA ratio than planktonic cells. Arrow indicates pCF10 band. **(J)** Comparison of pCF10 copy number in biofilm and planktonic cells exposed to cCF10 (0.1 ng/mL). Biofilm and planktonic cultures were induced with cCF10 for 60 minutes and sorted into "ON" and "OFF" populations. The sorted subpopulations were then analyzed for pCF10 copy number by qPCR. **This Figure is courtesy (Cook *et al.*, 2011).**

This demonstrates that biofilm cells have, on average, higher copy numbers of the conjugative plasmid pCF10 than their planktonic counterparts and also possess a greater heterogeneity in copy number compared to planktonic populations consistent with our mathematical model (Figure 6.2 C-E).

6.4 Discussion

In Chapters 4, the relative importance of the multiple intracellular mechanisms of regulation of expression of the pCF10 *prgQ* conjugation operon was discussed. Using mathematical modeling and quantitative analysis of transcript levels from the *prgQ* and *prgX* operons in response to cCF10 in planktonic cultures, with aid of various genetic constructs to analyze the specific contributions of the individual regulatory mechanisms to the system suggested, that the pheromone response system could function as a sensitive bistable genetic switch, and that disruption of any of the individual regulatory circuits of the system abolished switch behavior (Chatterjee *et al.*, 2011). However, the measurement of transcript levels by qRT-PCR represented the averages for the population neglecting possible heterogeneity in the response. A key finding of the switch model was the observation of a distinct "jump" in the production of Q_L RNA following exposure to pheromone concentrations in the range of 1-5 ng/ml. The findings in chapter 4 are consistent single-cell analysis shown in this chapter.

We employed a mathematical model to understand the effect of plasmid copy number on the conjugative response of the pCF10 system to inducer cCF10. The effect of increased plasmid copy number manifests itself through, increased number of operator

sites available for iCF10 bound PrgX tetramers and cCF10 bound PrgX dimers to bind and increased production of the iCF10 inhibitor peptide due to consequent increase in copies of Q_S and Q_L transcripts per donor cell. As a result, a higher dose of cCF10 is required to turn "ON" the genetic switch. Increasing the plasmid copy number widens the bistable region, suggesting that the switch behavior is more robust at higher plasmid copy numbers. Additionally, the wider the bistable region, a better separation between the "ON" and "OFF" populations is predicted, thus giving rise to observable bimodal distributions (Figure 6.3 A-D) as is observed in biofilm cells (Figure 6.3 B and D).

A pheromone-inducible GFP reporter construct was used for single cell level based expression analysis, to examine the effects of biofilm growth on the population dynamics of the response. The effects of biofilm growth on increasing both the average plasmid copy number (Figure 6.3 H) and copy number heterogeneity (Figure 6.3 J) are consistent with and validate the predictions of mathematical models described here (Figure 6.2) and in previous experiments (Chatterjee *et al.*, 2011). The lack of bimodal response in planktonic cultures might reflect the limitations of the detection system. It is likely that the low copy number of pCF10 in planktonic cells created a lower threshold for switch behavior to occur and also caused the cells to respond at a lower level blurring the distinction between "ON" and "OFF" (Figure 6.3 A and Figure 6.3 C).

Cells with higher copies of pCF10 would be poised to respond to cCF10 pheromone in a more controlled manner than cells with lower plasmid copy number. By requiring higher pheromone concentrations to turn ON conjugation, these cells can avoid

the expenditure of large amounts of energy involved in producing conjugative machinery. Biofilms are often seen as a type of multicellular community and allowing for conjugation to occur only in a subset of potential donor cells may be beneficial to the community as a whole. The physiological basis for the heterogeneity of copy number and pheromone response within the biofilm population is of great interest (Lemme *et al.*, 2011) and is currently under investigation. Our results provide insights into the control of conjugation at the single cell level, and highlight the importance of plasmid copy number distribution in biofilms.

6.5 Materials and Methods

6.5.1 MATHEMATICAL MODEL: A set of ordinary differential equations based on pCF10 genetic network were developed (Figure 1A). Steady state and dynamic simulations were performed in MATLAB (version 2008a MathWorks, Natick, MA). Steady state solutions of non-linear algebraic equations were obtained using MATLAB solve function. ODEs were integrated using function *ode23s* in MATLAB to obtain dynamic behavior.

6.5.2 EXPERIMENTAL PROCEDURES: For details see reference (Cook *et al.*, 2011).

CHAPTER 7

QUORUM SENSING VIA INHIBITOR SIGNALING MOLECULE ALLOWS REGULATION OF pCF10 CONJUGATION IN *E. faecalis*: A DYNAMIC RESPONSE TO INDUCTION

7.1 Summary

Pheromone-inducible conjugative transfer of the *Enterococcus faecalis* plasmid pCF10 is controlled by the antagonistic *prgQ* and *prgX* operons. The *prgQ* operon includes genes functionally involved in conjugation whereas the *prgX* operon encodes factors that negatively regulate conjugation. Recent modeling and experimental evidence showed that these operons reciprocally regulate each other via several mechanisms, including regulation of initiation of *prgQ* transcription, transcription interference and antisense interaction due to convergent transcription and confer a bistable switch behavior on the system. While previous studies have focused on pheromone induction of the *prgQ* operon, the present work addresses the mechanisms by which induced donor cells return to the uninduced state and the role of iCF10 in this process. Turning ON of *prgQ-prgX* genetic-switch induces expression of genes downstream of *prgQ*, encoding

various proteins involved in the transfer of the plasmid, in addition to production of an inhibitor of conjugation, thus giving rise to negative feedback loop. Both mathematical modeling and experimental analysis of dynamic response of pCF10 to induction indicate that this negative feed-back loop causes a brief "pulse-like" expression of the entire operon. Here, we describe the role of iCF10, the inhibitor signaling peptide for pCF10 based system, as quorum-sensing signal with the aim of turning-OFF conjugation at a population-wide level. Proteomic analysis showed dynamic expression of proteins encoded by the downstream genes, however, unlike transcript expression, these proteins were found to be comparatively more stable. Such an interplay of positive and negative feedback loops allowing the system to quickly transition between ON and OFF states may be critical both for the transfer of plasmid and survival of the donor cell.

7.2 Introduction

Horizontal gene transfer is one of the most common ways bacteria acquire antibiotic resistance and serves as the primary cause of the new generation of superbugs (Ochman *et al.,* 2000). Often the dissemination of antibiotic resistance and virulence determinants occurs via conjugative transfer of plasmid containing the gene of resistance from a donor to a recipient cell (Dunny, 2007a; Kurland *et al.*, 2003). In particular pheromone-inducible plasmid family of *Enterococcus faecalis* are one of the most efficient conjugation systems known (Dunny *et al.*, 1978; Dunny, 2007). Transfer of plasmid typically involves a signaling molecule based communication system between donor and recipient cells. In *Enterococcus faecalis*, a gram positive bacterium, recipient

cells release a chromosomally encoded hepta-peptide inducer molecule called cCF10 (LVTLVFV), which is transported into plasmid pCF10 carrying donor cells via a surface protein called PrgZ (Fixen *et al.*, 2007; Leonard *et al.*, 1996). On the other hand donor cells release an inhibitor molecule iCF10 (AITLIFI) that represses conjugation. Pheromone production is encoded within the core genome of *E. faecalis* and cells carrying pCF10 avoid self-induction by production of a plasmid-encoded inhibitor peptide called iCF10 (AITLIFI) (Nakayama *et al.*, 1994). This molecule acts as a competitive inhibitor of cCF10, and enables donor cells to distinguish between self (other donors) and non-self (recipients) (Buttaro *et al.*, 2000). The relative levels of iCF10 and cCF10 control conjugation as these compete for binding to the same repressor of conjugation PrgX (Figure 7.1 A) (Shi *et al.*, 2005).

Convergent transcription of operons *prgQ* and *prgX* in plasmid pCF10 regulates the conjugative transfer of plasmid pCF10 between *Enterococcus faecalis* donor and recipient cells (Chatterjee *et al.*, 2011; Dunny, 2007). The *prgQ* operon encodes conjugation machinery including expression of the downstream conjugation-related genes. The *prgX* operon encodes two negative regulators of conjugation, PrgX protein and Anti-Q, an abundant 104 nt RNA produced by processing of the 5' end of the *prgX* message (Bae *et al.*, 2002), which is antisense to *prgQ* transcripts. Additionally, transcriptional interference has been shown to result in generation of shorter truncated RNA from both P_Q and P_X promoters (denoted by Q_{AR} and X_{AR} RNA respectively, Figure

7.1 A) which have the potential to exert antisense interaction with full-length nascent P_X and P_Q transcripts respectively (Chatterjee *et al.*, 2011; Shearwin *et al.*, 2005)

PrgX negatively regulates the *prgQ* operon by binding as a dimer to two operator sites (XBS1 and XBS2; Figure 7.1) (Shi *et al.*, 2005). XBS2, the weaker binding site, overlaps the *prgQ* promoter P_Q, and its occupancy by PrgX reduces P_Q activity. Structural and genetic data support the model depicted in Figure 7.1 A whereby protein/protein interactions and DNA looping stabilize a repressing tetramer complex of iCF10-bound PrgX in uninduced cells (Dunny *et al.*, 1978; Shi *et al.*, 2005). During induction, cCF10 displacement of iCF10 and the resulting disruption of PrgX tetramers and the DNA loop ultimately reduces occupancy of XBS2 leading to increased P_Q activity. Even in the uninduced state, PrgX repression of P_Q is incomplete and a 380 nt transcript called Q_S is produced (Johnson *et al.*, 2010).

Recently, it has been demonstrated that interactions between Anti-Q and nascent *prgQ* transcripts can affect Q RNA folding and result in formation of a transcription terminator preventing read through into the downstream conjugation genes (at IRS1, Figure 7.1 A). The increased levels of *prgQ* transcription resulting from pheromone induction serves to titrate all of the Anti-Q RNA produced from the constitutively-expressed P_X, and leads to expression of conjugation genes by a post-transcriptional mechanism. Pheromone induction not only increases Q_S levels, and also results in expression of longer transcripts including the 530 nt Q_L, and even longer mRNAs extending through the operon, giving rise to expression of downstream conjugation-

related genes (Bensing *et al.*, 1997). The *prgQ* Orf, which encodes iCF10 is the only polypeptide-encoding gene within the Q_S region. A basal level of *prgQ* transcription and iCF10 synthesis is required to avoid self-induction (Buttaro *et al.*, 2000).

Previously, it has been shown that read through from the P_Q promoter gives rise to *prgB* transcription (Bensing *et al.*, 1996; Chung *et al.*, 1995). In fact, transcription from P_Q has been shown to extend up to 12 kb of inserted DNA indicating the role in inducing expression of downstream conjugation related genes (Bensing *et al.*, 1996; Chung and Dunny, 1992). The *prgQ* and *prgS* genes have been shown to function in a *cis* and orientation dependent manner to regulate expression of the downstream *prgB* gene (Chung *et al.*, 1995; Chung and Dunny, 1992), encoding the surface aggregation substance that produced by induced donor cells to bind to recipient cells and is exclusively expressed during induction (Olmsted *et al.*, 1991). Studies on the transcriptional response of pCF10 encoded genes have focused mainly on the turning ON response upstream of *prgB* gene (Bae *et al.*, 2002; Chatterjee *et al.*, 2011; Johnson *et al.*, 2010). However, an unexplored aspect of this switch behavior is the mechanism by which induced donor cells return back to the uninduced state.

In the present study, we demonstrate that downstream conjugation related pCF10 transcripts exhibit a pulse of expression in response to induction with nearly 2-3 orders increase in expression within 30 minutes, followed by a significant decrease in expression within 2 hours post-induction, observed for both physiological levels as well as very high levels of inducer cCF10. Previous transcriptome study using a mini-

microarray based on 57-Orfs on pCF10 showed a similar "pulse-like" dynamic expression of downstream pCF10 genes, however, induction was performed only at moderate levels of pheromone concentration (Hirt *et al.*, 2005). Our results indicate that the inhibitor iCF10 functions as a natural quorum sensing molecule for donor cells to sense their own population and inhibit conjugation. Both mathematical modeling and experiments indicate pCF10 conjugation machinery is intrinsically programmed to turn-OFF conjugation, and turning-ON of conjugation prompts turning-OFF via iCF10 quorum. Remarkably, protein expression from pCF10 genes downstream from *prgQ* continued even after the transcript expression subsided.

7.3 Results

7.3.1 Mathematical model of dynamics of pCF10-based induction of conjugation

A mathematical model to study the dynamics of pCF10 based conjugation system consisting of a set of 12 differential equations shown in Table 7.1 was employed based on a previous model reported in (Chatterjee *et al.*, 2011) (Figure 7.1 A). The parameters used in the model are summarized in Table 7.2. Transcription from P_Q depends on whether it is in the depressed state (unlooped state) or repressed state (looped state), whereas expression from P_X is constitutive (Chatterjee *et al.*, 2011). However, due to transcriptional interference exerted by promoter P_Q on P_X and vice versa, the apparent transcription rate from both P_Q and P_X promoters depend on the concentration of pCF10

plasmid in the looped vs unlooped state (Chatterjee *et al.*, 2011). After considering transcriptional interference, under uninduced /induced state the rate constants for transcription of nascent P_Q RNA ($K_{P_Q,R}$ and $K_{P_Q,D}$), nascent P_X RNA ($K_{P_X,R}$ and $K_{P_X,D}$), Q_{AR} RNA ($K_{Q_{AR},R}$ and $K_{Q_{AR},D}$) and X_{AR} RNA ($K_{X_{AR},R}$ and $K_{X_{AR},D}$) are summarized in Table 7.2 (Chatterjee *et al.*, 2011).

Truncated P_Q transcripts produced as a result of transcriptional interference that are capable of exerting antisense effects on X RNA are assumed to follow second order irreversible kinetics as shown in Equation 7.5-7.6. Similarly, truncated P_X transcripts that are capable of exerting antisense effects on nascent P_Q RNA and Anti-Q RNA are collectively denoted as X_{AR} and are considered in Equations 7.2-7.4 (Chatterjee *et al.*, 2011). Anti-Q RNA is assumed to have 1:1 stoichiometry with full length X RNA, based on the fact that X transcripts are processed at the 5' end to generate Anti-Q RNA (Bae *et al.*, 2000; Bae *et al.*, 2002). Nascent transcripts from P_Q promoter that interact with X_{AR}, terminate to form Q_S RNA, whereas those that escape interaction with X_{AR} form Q_L RNA (Chatterjee *et al.,* 2011). The net rate of production of Q_S and Q_L RNA is given by the first term in Equations 7.2 and 7.3 respectively.

The net transcription rate of transcripts Q_S, Q_L and Q_{AR} from P_Q and X and X_{AR} from P_X is denoted by the sum of transcription rate in the uninduced and induced states weighted by concentration of occupied [O] to unoccupied operator sites ([N]-[O]) respectively as shown in Equations 7.1-7.6. The induced or uninduced state of P_Q is determined by the formation of DNA loop at operator sites XBS 1 and XBS 2 (Figure 7.1

A). The binding of PrgX-iCF10 tetramers and PrgX-cCF10 dimers to XBS 1 and 2 is treated as a rapid equilibrium process (Shi *et al.*, 2005). Consequently, the concentration of pCF10 DNA in the looped state ([O]) is given by Equation 1 and has been derived elsewhere (Chatterjee *et al.*, 2011).

To model the process of pCF10 conjugation, the ODE model described in (Chatterjee *et al.*, 2011) was extended to include mass-action kinetics based equations for three additional transcripts on pCF10, namely, *prgB, pcfC* and *pcfG* as shown in Equations 7.7-7.9. These genes were considered in the model as these are placed at various locations within the set of downstream conjugation-related genes present on pCF10 (Hirt *et al.*, 2005). While Q_L RNA is an indicator for the onset of conjugation, *prgB* gene which is approximately 4.7 kb downstream from *prgQ* start site encodes the surface aggregation substance and corresponds to an intermediate step during conjugation (Bensing and Dunny, 1997). The genes *pcfC* and *pcfG* are approximately 20 kb and 25 kb downstream of *prgQ* start site respectively (Hirt *et al.*, 2005), encode for Type IV secretion system coupling protein and *relaxase* protein respectively (Chen *et al.*, 2008), which participate in the final process of transfer of plasmid from the donor to the recipient cell (Chen *et al.*, 2007; Chen *et al.*, 2008).

Transcription rate of *prgB*, *pcfC* and *pcfG* RNA is assumed to be proportional to Q_L RNA as shown in Equations 7.7-7.9, in addition to a basal transcription rate denoted by the first term $K_{basal_{prgB}}$, $K_{basal_{pcfC}}$ and $K_{basal_{pcfG}}$ in Equations 7.7, 7.8 and 7.9 respectively. A delay in transcription of *prgB, pcfC* and *pcfG* gene post-transcription of Q_L RNA is

assumed to occur due to time taken by RNA polymerase to read through approximately 27.1 Kbp the *prgQ* operon consisting of the downstream conjugation genes. The Orf of *prgB*, *pcfC* and *pcfG* are located 8078, 21448 and 26251 bp respectively downstream from the 3' end Q_L RNA (Hirt *et al.*, 2005) (Figure 7.1 B). Assuming that the average velocity of RNA polymerase in presence of pausing is 50 bp/s (Neuman *et al.*, 2003), a time-delay of 2.7, 7.1 and 8.7 minutes in transcription post expression of Q_L RNA is estimated for *prgB*, *pcfC* and *pcfG* genes respectively (Equations 7.7-7.9). For the various RNA species, degradation and dilution due cell growth is considered to be first order as shown in Equations 7.2-7.9.

The dynamics of extracellular iCF10 (i), intracellular iCF10 (I), extracellular cCF10 (c) and intracellular cCF10 (C) is shown in Equations 7.10-7.13. The transport of signaling molecules cCF10 and iCF10 across the cell membrane is modeled as a first order reaction (Equation 7.10-7.13). Generation of extracellular iCF10 (i) is modeled as first order with respect to both Q_S and Q_L denoted by the first term in Equation 7.10. The amplification of extracellular iCF10 due to quorum sensing effect of donor population is denoted by the multiplication factor 'Q' in Equation 7.11. Degradation and dilution due cell growth is considered to be first order as shown in Equations 7.10-7.13. Dilution due to growth is not considered for extracellular iCF10 and cCF10, however, extracellular degradation is considered in Equation 7.11 and 7.12 respectively.

The Equations 7.1-7.12 were solved for steady state for a fixed concentration of extracellular cCF10 to estimate the steady response of *prgQ-prgX* switch to induction,

whereas, Equations 7.1-7.13 were solved for estimating dynamic response of donor cells to induction for a given impulse of extracellular cCF10 concentration. The bistable steady state response of Q_L to inducer has been previously shown to be theoretically robust and spanning over a wide range of parameter space (Chatterjee *et al.*, 2011). The rate constants for the dynamics simulations shown in Table 7.2 correspond to a bistable steady state response of Q_L to induction with cCF10.

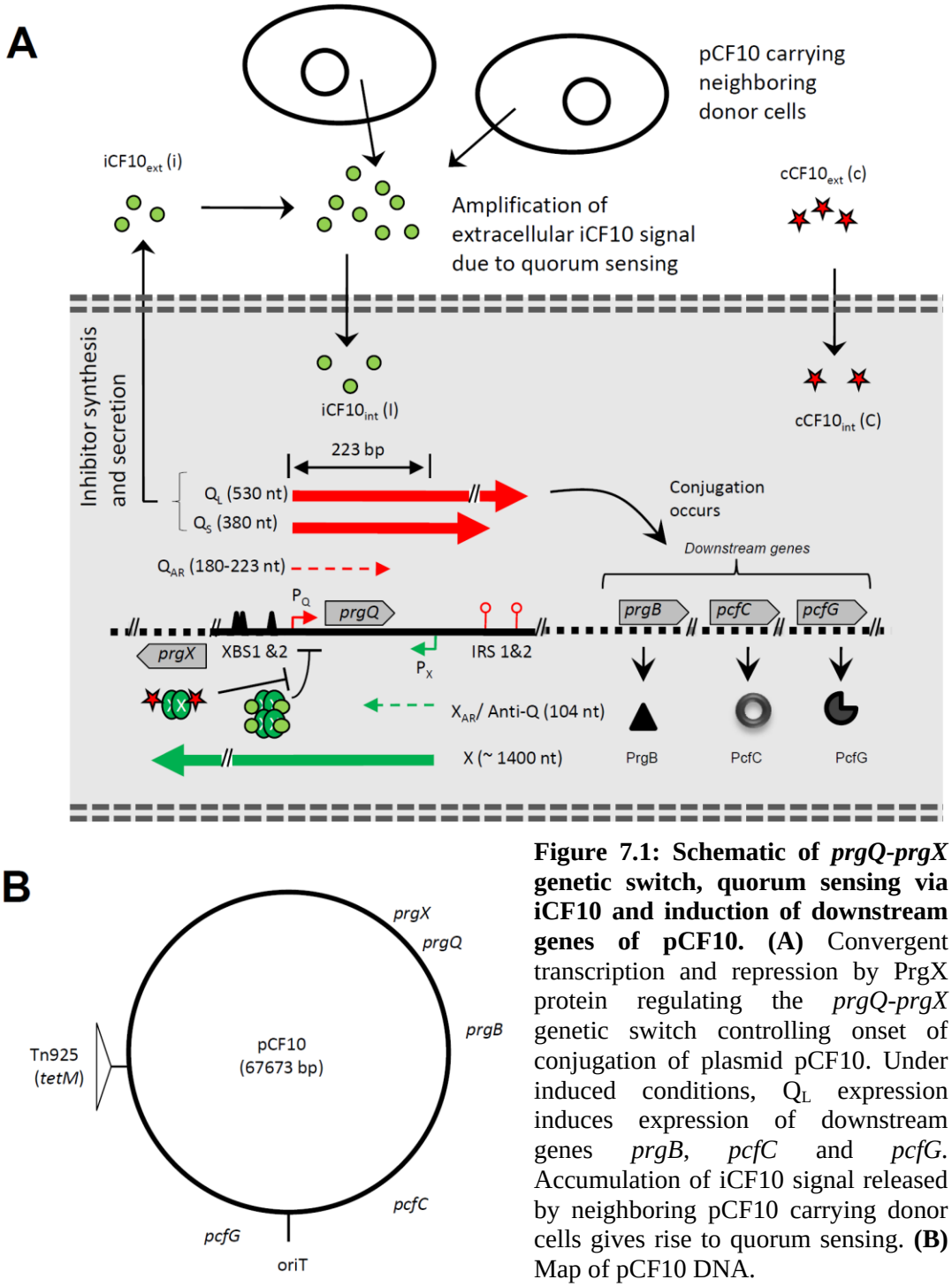


Figure 7.1: Schematic of *prgQ-prgX* genetic switch, quorum sensing via iCF10 and induction of downstream genes of pCF10. (A) Convergent transcription and repression by PrgX protein regulating the *prgQ-prgX* genetic switch controlling onset of conjugation of plasmid pCF10. Under induced conditions, Q_L expression induces expression of downstream genes *prgB*, *pcfC* and *pcfG*. Accumulation of iCF10 signal released by neighboring pCF10 carrying donor cells gives rise to quorum sensing. **(B)** Map of pCF10 DNA.

Table 7.1: List of ODEs for the mathematical model

S.No.	Differential Equations
7.1	$[O] = \frac{N \cdot [I]^4}{[I]^4 + K_B \cdot [C]^4}$
7.2	$\frac{d[Q_S]}{dt} = K_{Q,R}[O] + K_{Q,D}([N]-[O]) \left(\frac{K_{Q-X_{AR}} \cdot [X_{AR}]}{1 + K_{Q-X_{AR}} \cdot [X_{AR}]} \right) - (\lambda_{Q_S} + \mu)[Q_s]$
7.3	$\frac{d[Q_L]}{dt} = K_{Q,R}[O] + K_{Q,D}([N]-[O]) \left(\frac{1}{1 + K_{Q-X_{AR}} \cdot [X_{AR}]} \right) - (\lambda_{Q_L} + \mu)\ Q_L$
7.4	$\frac{d[X_{AR}]}{dt} = K_{X_{AR},R}[O] + K_{X_{AR},D}([N]-[O]) - K_{AR}[X_{AR}][Q_{AR}] - [K_{Q,R}[O] + K_{Q,D}([N]-[O])] \left(\frac{K_{Q-X_{AR}} \cdot [X_{AR}]}{1 + K_{Q-X_{AR}} \cdot [X_{AR}]} \right) - (\lambda_{X_{AR}} + \mu)[X_{AR}]$
7.5	$\frac{d[X]}{dt} = K_{X,R}[O] + K_{X,D}([N]-[O]) - K_{AR}[X] \cdot [Q_{AR}] - (\lambda_X + \mu)[X]$
7.6	$\frac{d[Q_{AR}]}{dt} = [K_{Q_{AR},R}[O] + K_{Q_{AR},D}([N]-[O])] - K_{AR}[X_{AR}] \cdot [Q_{AR}] - (\lambda_{Q_{AR}} + \mu)[Q_{AR}]$
7.7	$\frac{d[prgB]}{dt} = K_{basal_{prgB}} + K_{prgB}[Q_L]\tau(t - t_{prgB}) - (\lambda_{prgB} + \mu)[prgB]$
7.8	$\frac{d[pcfC]}{dt} = K_{basal_{pcfC}} + K_{pcfC}[Q_L]\tau(t - t_{pcfC}) - (\lambda_{pcfC} + \mu)[pcfC]$
7.9	$\frac{d[pcfG]}{dt} = K_{basal_{pcfG}} + K_{pcfG}[Q_L]\tau(t - t_{pcfG}) - (\lambda_{pcfG} + \mu)[pcfG]$
7.10	$\frac{d[I]}{dt} = K_{T_i}[i].V_{iCF10} - (\lambda_I + \mu)[I]$
7.11	$\frac{d[i]}{dt} = K_i([Q_s] + [Q_L]) \cdot V_{iCF10}.Q - K_{T_i} \cdot [i]/V_{iCF10} - \lambda_i \cdot [i]$
7.12	$\frac{d[C]}{dt} = K_{T_c}[c].V_{cCF10} - (\lambda_C + \mu)[C]$
7.13	$\frac{d[c]}{dt} = -K_{T_c}[c]/V_{cCF10} - \lambda_c \cdot [c]$

Table 7.2: Parameters used in dynamic simulations

Parameter	Description	Value used in Ch 4-5	Values used in Ch 7	Units
μ	Specific growth rate of donor cells	2.58×10^{-4}	2.58×10^{-4}	s^{-1}
$K_{Q,R}$	Transcription rate constant of Q in uninduced state	7.50×10^{-3}	7.50×10^{-3}	s^{-1}
$K_{Q,D}$	Transcription rate constant of Q in induced state	8.87×10^{-2}	8.87×10^{-2}	s^{-1}
$K_{X_{AR},R}$	Transcription rate constant of X_{AR} in uninduced state	1.02×10^{-2}	1.02×10^{-2}	s^{-1}
$K_{X_{AR},D}$	Transcription rate constant of X_{AR} in induced state	1.20×10^{-3}	1.20×10^{-3}	s^{-1}
$K_{X,R}$	Transcription rate constant of X in uninduced state	1.2×10^{-3}	1.2×10^{-3}	s^{-1}
$K_{X,D}$	Transcription rate constant of X in induced state	1.20×10^{-3}	1.20×10^{-3}	s^{-1}
$K_{Q_{AR},R}$	Transcription rate constant of Q_{AR} in uninduced state	1.80×10^{-3}	1.80×10^{-3}	s^{-1}
$K_{Q_{AR},D}$	Transcription rate constant of Q_{AR} in induced state	1.53×10^{-3}	1.53×10^{-3}	s^{-1}
K_i	Generation rate of extracellular iCF10	1×10^{-3}	2×10^{-1}	s^{-1}
V_{iCF10} *or* V_{cCF10}	Volume conversion factor for iCF10 and cCF10	1	1	-
Q	Amplification factor due to quorum sensing of iCF10	-	100	-
K_{T_i}	Transport rate constant of iCF10	1×10^{-3}	1.5×10^{-5}	s^{-1}
K_{T_C}	Transport rate constant of cCF10	1×10^{-3}	1×10^{-2}	s^{-1}
K_{AR}	Rate constant of interaction between X RNA and Q_{AR} RNA	1×10^{-3}	1×10^{-3}	$(nM)^{-1}$. s
$K_{P_Q - X_{AR}}$	Equilibrium constant of Q and X_{AR} interaction	4.43	4.43	$(nM)^{-1}$
K_B	Equilibrium constant of DNA binding reaction	1×10^{8}	1×10^{8}	-
λ_{Q_S}	Degradation rate of Q_S RNA	1×10^{-3}	1×10^{-3}	s^{-1}
λ_{Q_L}	Degradation rate of Q_L RNA	1×10^{-1}	1×10^{-1}	s^{-1}
$\lambda_{X_{AR}}$	Degradation rate of X_{AR} RNA	1.36×10^{-4}	1.36×10^{-4}	s^{-1}
$\lambda_{Q_{AR}}$	Degradation rate of Q_{AR} RNA	1×10^{-3}	1×10^{-3}	s^{-1}
λ_X	Degradation rate of X RNA	1×10^{-4}	1×10^{-4}	s^{-1}
λ_i	Degradation rate of extracellular iCF10	1×10^{-6}	1×10^{-6}	s^{-1}
λ_c	Degradation rate of extracellular cCF10	1×10^{-6}	1×10^{-6}	s^{-1}
λ_I	Degradation rate of intracellular iCF10	1×10^{-6}	1×10^{-6}	s^{-1}
λ_C	Degradation rate of intracellular cCF10	1×10^{-6}	1×10^{-6}	s^{-1}
$k_{basal_{prgB}}$	Basal transcription rate of *prgB* RNA	-	3×10^{-4}	s^{-1}
$k_{basal_{pcfC}}$	Basal transcription rate of *pcfC* RNA	-	3×10^{-4}	s^{-1}
$k_{basal_{pcfG}}$	Basal transcription rate of *pcfG* RNA	-	3×10^{-4}	s^{-1}
K_{prgB}	Transcription rate of *prgB* RNA	-	1×10^{-2}	s^{-1}
K_{pcfC}	Transcription rate of *pcfC* RNA	-	1×10^{-2}	s^{-1}
K_{pcfG}	Transcription rate of *pcfG* RNA	-	1×10^{-2}	s^{-1}
t_{prgB}	Time-delay in transcription of *prgB* RNA	-	2.7	Min
t_{pcfC}	Time-delay in transcription of *pcfC* RNA	-	7.1	Min
t_{pcfG}	Time-delay in transcription of *pcfG* RNA	-	8.6	Min
λ_{prgB} λ_{pcfC} λ_{pcfG}	Degradation rate of *prgB, pcfC* and *pcfG* RNA	-	1×10^{-3}	s^{-1}

7.3.2 The iCF10 quorum results in a negative feedback to turn-OFF prgQ-prgX genetic switch

The objective of the mathematical model was to identify the conditions under which the system demonstrates a turning-OFF response (Figure 7.2). The turn-ON of the system is characterized by increased expression level of Q_L RNA, which further leads to elevated expression of downstream genes such as *prgB, pcfC* and *pcfG* in the model. System is considered to be in OFF state when expression levels of various conjugation-related transcripts return to the corresponding to uninduced state.

As expected the bistable steady state response of Q_L expression to cCF10 demonstrates a shift towards the right for increased levels of iCF10 quorum (denoted by factor Q), indicating the need for higher levels of cCF10 concentration when donor population increases. The dynamic response of donor cells to induction with different levels of cCF10 obtained by solving Equations 7.2-7.13 is shown in Figure 7.2 B-C for different levels of iCF10 quorum. The simulation parameters (Table 7.2) used were fitted to experimental data shown in Figure 7.5-7.7 and corresponds to a bistable steady state behavior (Figure 7.2 A, case Q=100 and B). Mathematical model predicts that induction of pCF10 results in a brief surge of transcription along the *prgQ* operon post-induction only at high values of iCF10 quorum (Q=100) as shown in Table 2 (Figure 7.2 B). When the iCF10 quorum was decreased by an order (Q=10) such a pulse behavior is lost at higher levels of cCF10 (Figure 7.2 C). Under such conditions, Q_L transcript level will increase and stay at high level for long durations of time, presenting a high energy cost

for the donor cell due to metabolic burden imposed for continuously maintaining the cell at an ON state.

The model predicts that the system transitions between OFF to ON state and concentrations greater than 0.5 ng/mL of cCF10 (for Q>10) consistent with previous data (Chatterjee *et al.*, 2011; Cook *et al.*, 2011) (Figure 7.2 A-C). Q_L expression levels are predicted to increase by a couple of orders and decrease very quickly within a time scale of 1-2 hours, thus demonstrating a pulse of transcription. A similar pulse of transcription is predicted to be propagated through downstream conjugation related genes (Figure 7.3). The model predicts that two main factors are needed for this pulse behavior, namely, convergent transcription of *prgQ-prgX* gene pair which allows for a bistable switch behavior and high iCF10 quorum. These criteria are also essential for prediction of bistable steady state response to induction with cCF10 (Chatterjee *et al.*, 2011).

Interestingly, comparison of a condition with bistability (Figure 7.4 A), vs. condition lacking a bistable (Figure 7.4 B), indicates that bistability results in a faster OFF response. The conditions without bistability correspond to RC/AR^- case where antisense RNA interaction is removed from the model as discussed in (Chatterjee *et al.*, 2011). The corresponding dynamic response of Q_L RNA for the two conditions is shown respectively in Figure 7.4 C-D. The most efficient turning OFF response is predicted for the case with bistability, which is also apparent from the corresponding Q_L transcription rate (Figure 7.4 E-F).

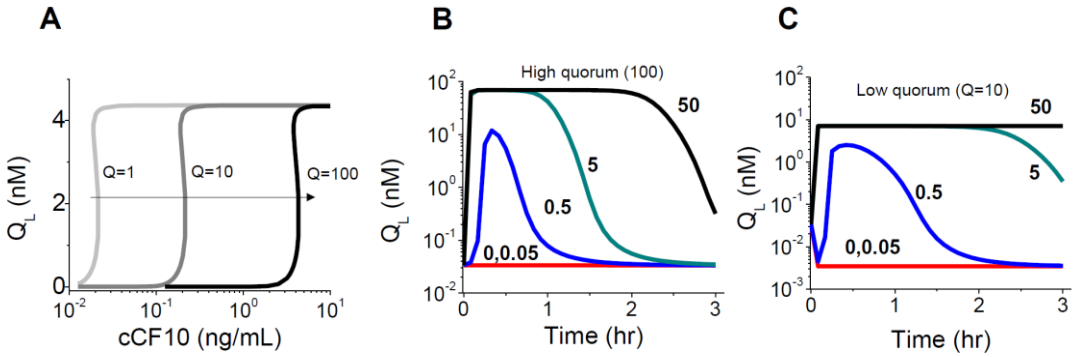


Figure 7.2: Dynamic and steady state expression profile of Q_L RNA in response to induction predicted by the mathematical model. (A) Steady state bistable curve for different iCF10 secretion rates. **(B)** High iCF10 secretion rate constant (10X). **(C)** Low iCF10 secretion rate constant (1X)

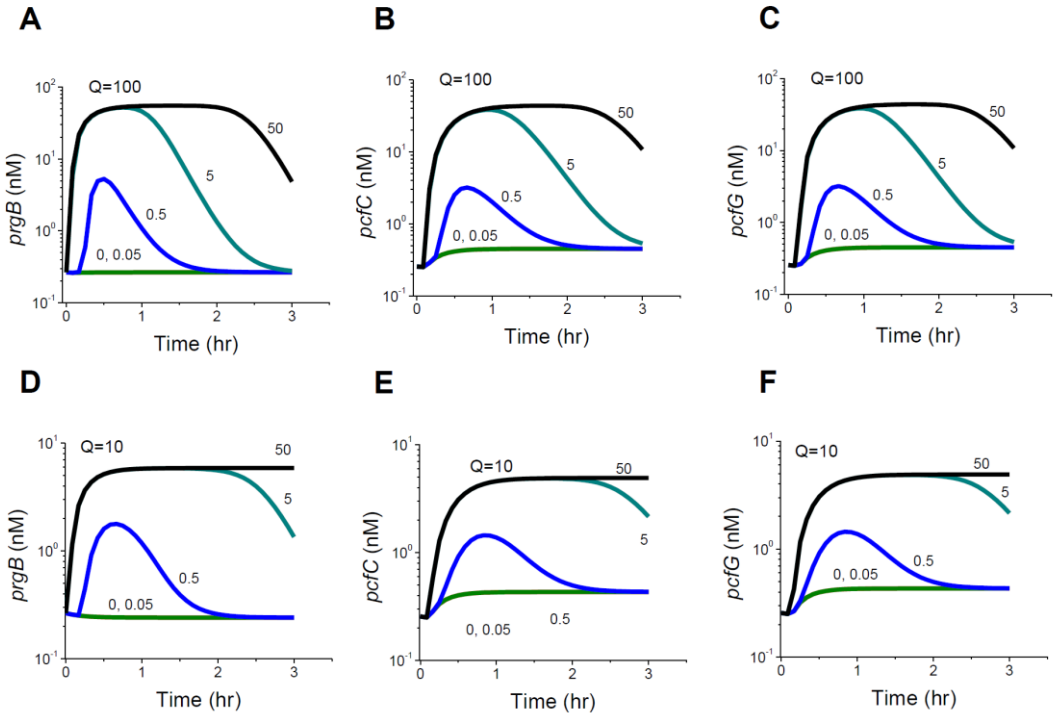


Figure 7.3: Dynamic expression profile of *prgB*, *pcfC* and *pcfG* RNA for different levels of quorum-sensing factor in response to induction. (A-C) Response to induction with different levels of cCF10 to high iCF10 quorum (Q=100). **(D-F)** Response to induction with different levels of cCF10 to low iCF10 quorum (Q=10).

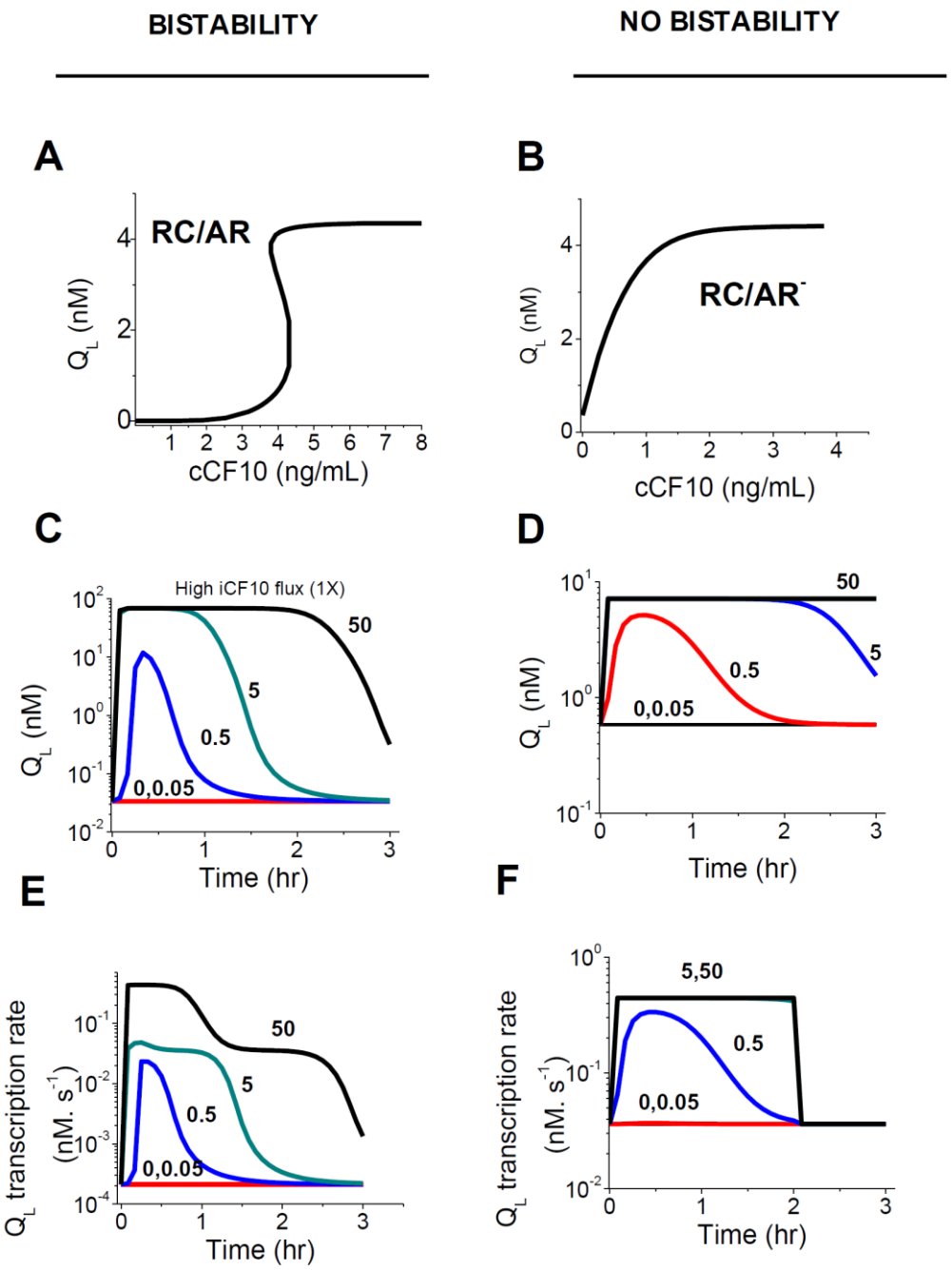


Figure 7.4: Steady state and transient behavior of Q_L expression and Q_L transcription rate in response to induction. (A-B) Steady state response to induction in presence (A) and absence of bistability at iCF10 quorum (Q=100) (B). **(C-D)** Dynamic response for corresponding conditions shown in A-B. **(E-F)** Transcription rate of Q_L RNA with time in response to induction for the conditions shown in A-D.

7.3.4 Transcription pattern of pCF10 post-induction: a pulse of transcript expression

We next looked at the dynamic response of transcripts from the pCF10 plasmid involved in conjugation. Overnight cultures of pCF10 carrying donor cells in host strain OG1Sp were diluted 1:10 and induced with different concentrations of cCF10 varying between 0 to 50 ng/mL. We chose very low concentration of cCF10 such as 0.05 ng/mL, which is similar to those found in the supernatant of a recipient culture (Nakayama *et al.*, 1994), as well as very high concentration of cCF10, up to 1000 times that of the physiological concentration (50 ng/mL). Samples were collected from exponentially growing cells within first two hours of induction (Figure 7.5 A). The dynamics of expression of transcripts Q_L, *prgB*, *pcfC* and *pcfG* was measured at various time points using qPCR. Time course data showing dynamics of induction with different concentrations of cCF10 confirm presence of a transcriptional pulse (Figure 7.5 B-D), consistent with the predictions of the mathematical model (Figure 7.2 B-C, Figure 7.3 A-C).

Pulse of transcript expression was observed for induction with both very low (0.05 ng/mL) as well high concentrations (50 ng/mL) of cCF10. Q_L expression level increased nearly 2 orders of magnitude by 15 minutes post-induction, reaching peak expression by 30 minutes (Figure 7.5 B). Thereafter Q_L expression levels decreased to that corresponding to OFF state within 1 hour post-induction at low to moderate level of induction (0.05 and 0.5 ng/mL of cCF10). At high levels of induction (5 and 50 ng/mL of

cCF10), Q_L expression levels showed higher at 15 minutes than that observed at 0.05 ng/mL (Figure 7.5 B), most-likely due to increased inducer levels, thereafter Q_L expression level decreased almost immediately. Though Q_L expression did not decrease back to initial state at high induction, a decreasing trend in transcript expression was observed, pointing towards a switching-OFF phenotype even at this high level of induction.

A similar transcript expression profile was observed for transcripts *prgB*, *pcfC* and *pcfG* corresponding to the conjugation-related genes on pCF10 (Figure 7.5 C-D), with transcript expression peaking at 30 minutes for the entire range of cCF10 concentrations used (0.05-5 ng/mL). Similar to Q_L expression, *prgB*, *pcfC* and *pcfG* transcript expression decreased to initial expression levels by two hours post-induction for low to moderate cCF10 levels (Figure 7.6 C-E, 0.05 and 0.5 ng/mL of cCF10). At high levels of induction (5 and 50 ng/mL of cCF10), transcript expression nearly reached the initial OFF state. Approximately 2-3 orders of magnitude increase and decrease in expression was observed within 2 hours post-induction across all the four transcripts (Figure 7.2 B-D) for very low as well as very high levels of cCF10.

Comparing the transcript expression profiles of Q_L, *prgB*, *pcfC* and *pcfG* RNA shows that pulse of transcription for *prgB*, *pcfC* and *pcfG* RNA lags that of Q_L RNA. A shorter time-course experiment at high levels of induction (50 ng/mL of cCF10) showed that upstream genes (Q_L) are expressed earlier than downstream genes (*prgB*, *pcfC*, *pcfG*) (Figure 7.6 A-D). Surprisingly, Q_L expression increased significantly by as early as 5

minutes post-induction (Figure 7.6 A), which was not observed for *prgB*, *pcfC* and *pcfG* RNA. While Q_L transcript expression reached peak levels by 15 minutes post-induction, *prgB*, *pcfC* and *pcfG* reached peak expression by 30 minutes (Figure 7.6 B-D). This supports the model assumption of delayed expression of downstream transcripts.

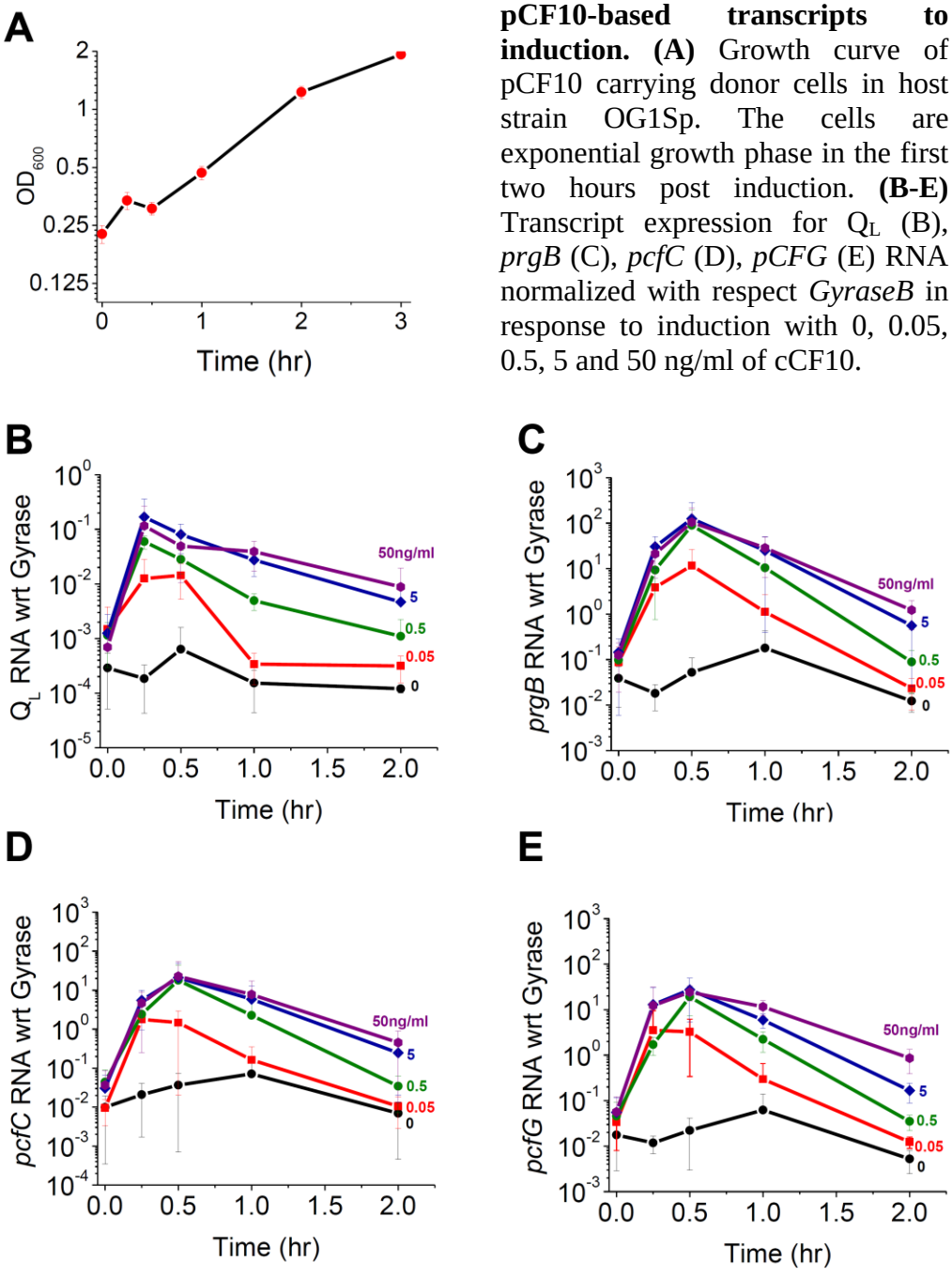


Figure 7.5: Dynamic responses of pCF10-based transcripts to induction. (A) Growth curve of pCF10 carrying donor cells in host strain OG1Sp. The cells are exponential growth phase in the first two hours post induction. **(B-E)** Transcript expression for Q_L (B), *prgB* (C), *pcfC* (D), *pCFG* (E) RNA normalized with respect *GyraseB* in response to induction with 0, 0.05, 0.5, 5 and 50 ng/ml of cCF10.

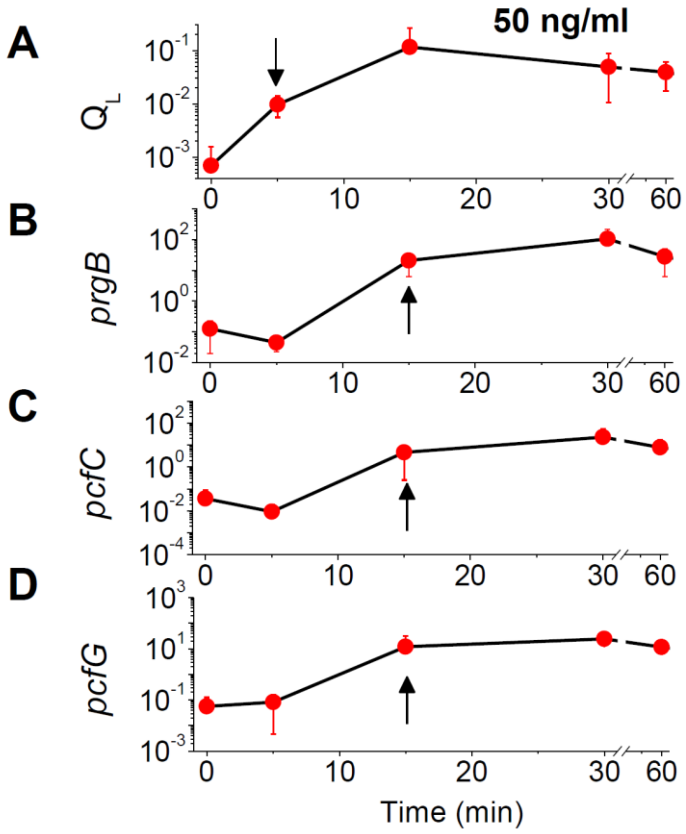


Figure 7.6:
Early response of pCF10-based transcripts to induction at high levels of induction (50 ng/ml).

7.3.5 The prgQ-prgX gene pair is sufficient to cause a pulse of transcription

To evaluate whether *prgQ-prgX* gene pair and/or additional pCF10-based factors influence the pulse-like dynamic response of pCF10, we measured expression from pBK2 plasmid in host strain OG1Sp in response to induction with cCF10. Plasmid pBK2, contains only the *prgQ-prgX* gene pair, with *lacZ* gene fused downstream of IRS1 and serving as a reporter for Q_L RNA (Figure 7.7 A), thus excluding other factors present on pCF10 which could influence the dynamic response to induction. pBK2 carrying cells

were induced with high levels of cCF10 (50 ng/mL). LacZ expression was measured using QPCR to estimate the dynamic response of *prgQ-prgX* gene pair alone to induction.

A pulse of *lacZ* transcript expression was observed, thus indicating that the *prgQ-prgX* locus alone is sufficient in attributing such transient dynamics (Figure 7.7 A). However in this case, after 2 hours of induction at 50 ng/mL of cCF10, only a 4 fold decrease in *lacZ* expression was observed from the peak expression levels (Figure 7.7 A). On the other hand, after 2 hours of induction of the wildtype pCF10 system (in OG1Sp) at 50 ng/mL of cCF10, Q_L transcript demonstrated nearly 2 orders of magnitude decrease from the peak level of expression (seen at 30 min to 1 hr). This indicates that the potentially shorter half-life of Q_L RNA is essential for the system to completely shut-down and return to ground state in a relatively short period of time.

7.3.6 Removal of iCF10 flux abolishes transcriptional pulse

A key prediction of the mathematical model was that the pulse of expression of Q_L RNA occurs at high iCF10 quorum. To test this hypothesis, we evaluated the dynamic response to induction of pCF10 carrying donor cells in absence of iCF10 in host strain OG1RF with an in frame deletion in the Eep Orf (ΔEep) producing a non-functional Eep protein (described in (Kristich *et al.*, 2007)). Eep protein is responsible for proteolytic cleavage of pre-iCF10 and pre-cCF10 into mature iCF10 and cCF10 peptides respectively (Chandler and Dunny, 2008). As a result, the ΔEep mutant is deficient in production of mature iCF10 and cCF10. pCF10 carrying ΔEep mutants, were induced with high levels of cCF10 (50 ng/mL) and allowed to grow in absence of iCF10 (Figure

7.7 B-C). Interestingly, in absence of iCF10 influx, the transcription-pulse behavior was not observed for both Q_L and *prgB* transcripts (Figure 7.7 B-C). Transcription-pulse was restored when the same culture was supplemented with 10 fold higher level of iCF10 (500 ng/ml) added at 60 minutes post-induction (Figure 7.7 B-C). A similar response was observed for both Q_L RNA as well as *prgB* RNA, though Q_L RNA was expressed at lower level and restored to initial levels (Figure 7.7 B) compared to the *prgB* transcript (Figure 7.7 C). This indicates that the most important factor in controlling the transcriptional pulse is the negative feed-back from iCF10. This evident from the fact that even though the convergent transcription operates in the *prgQ-prgX* gene-pair, still in the absence of iCF10, the pulse response is not observed. Hence, the negative feedback from iCF10 is essential for the system to turn back off, consistent with the model prediction.

7.3.7 Dynamics of protein expression from plasmid pCF10 in response to induction

We next evaluated the dynamic response of proteins expressed from pCF10 plasmid in response to induction. pCF10 carrying donor cells were induced with high levels of inducer (50 ng/mL of cCF10). Since dynamic expression of Q_L, *prgB, pcfC* and *pcfG* transcript occurred within two hours post-induction (Figure 7.5-7.7), samples for proteomic analysis were collected at similar time points, i.e., at 0, 30, 60 and 120 minutes post-induction (see Materials and methods). Whole cell proteome was labeled using 8-plex-ITRAQ labeling method (LC-MALDI-TOF) from two biological replicates. A total

of 397 proteins were identified with a 5% False Discovery rate, out of which 16 were pCF10 based.

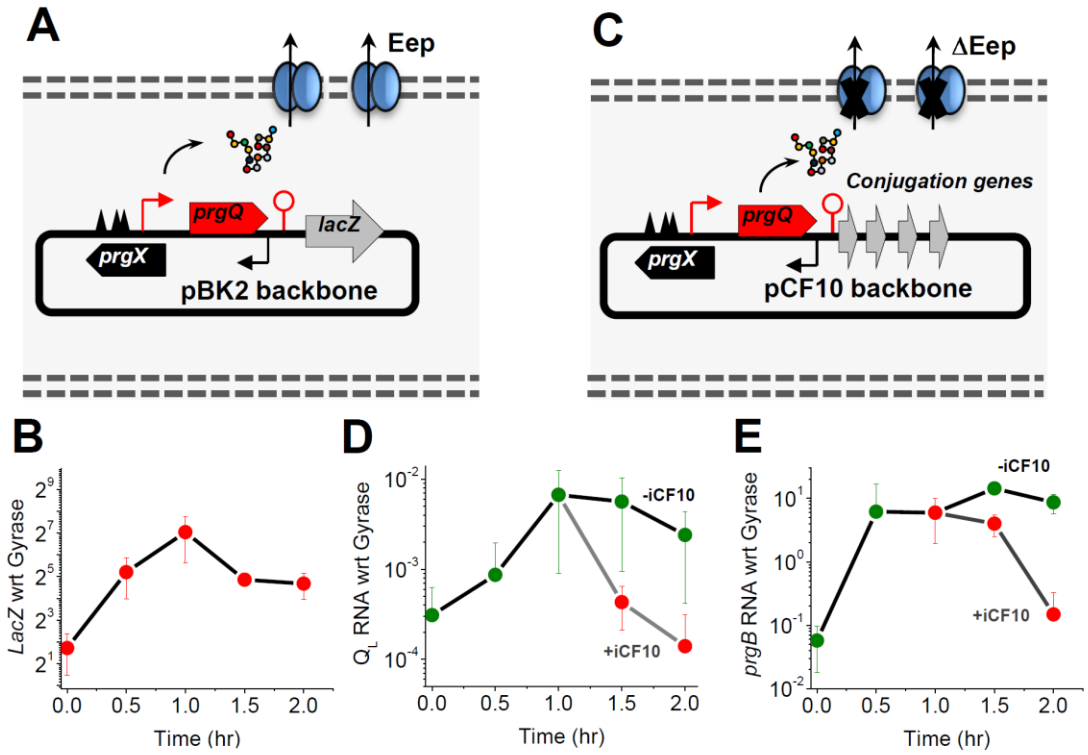


Figure 7.7: Transcription-pulse requires *prgQ-prgX* gene-pair and iCF10 flux. **(A)** Schematic of plasmid pBK2. **(B)** Schematic of OG1RF-ΔEep+pCF10 construct. **(C)** Dynamic response of pBK2 plasmid in host strain OG1Sp at high levels of induction with 50 ng/ml of cCF10. *LacZ* expression was measured using QPCR and is normalized to *GyraseB*. **(D-E)** Dynamic response of wildtype pCF10 plasmid in host strain OG1RF-ΔEep mutant, which expresses a non-functional Eep protein incapable of producing mature iCF10 and cCF10 peptides. The dynamic response of Q_L RNA (D) and *prgB* RNA (E) to induction with cCF10 (50 ng/ml) in absence of iCF10 is indicated by the black curve and in presence of iCF10 (500 ng/ml) added at 60 minutes post-induction is indicated by the grey curve.

The dynamic expression profile of proteins from pCF10 post-induction is shown in Figure 7.8. Out of the 16 proteins identified from pCF10, 13 of them are expressed

from genes downstream of *prgQ-prgX* gene pair (Figure 7.8 A-M), and three from genes upstream of *prgQ-prgX* gene pair (Figure 7.8 N-P). Interestingly, among the pCF10 based proteins, dynamic expression was observed only for those encoded from downstream genes, whereas, for the proteins upstream the expression profile was more or less constant. Interestingly, for the pCF10 based proteins that were detected by the mass-spectrometry, the corresponding genes were found to be physically located closer to each other on pCF10 plasmid. Functionally, these proteins can be classified into three functional classes, namely, (I) Surface proteins, (II) Mating channel proteins and (III) DNA processing proteins.

PrgA, PrgB and PrgU encode for surface proteins, while PrgA is a surface exclusion protein which allows the donor cell to differentiate between self and non-self (Bryan *et al.*, 2000; Dunny *et al.*, 1985; Hirt *et al.*, 2005), PrgB is a surface aggregation protein which allows the donor to anchor to the recipient cell (Chuang *et al.*, 2009). The function of PrgU has not yet known, but this short 118 amino acid protein shares 98% protein-sequence homology with *orf3* of plasmid pAD1 plasmid in *Enterococcus faecalis* (Hirt *et al.*, 2005). All three proteins demonstrated 2-4 fold increase in expression in response to induction within 30-60 minutes post induction between the two biological replicates.

The proteins PrgD, PrgE, PrgJ, PrgM, PrgL, PcfA and PcfB encode for proteins that form the mating channel between the donor and the recipient cell (Chen *et al.*, 2008). Again, expression increased significantly (2-4 fold) within 30-60 minutes post-induction

for all these proteins (Figure 7.8 D-J). Most of these proteins have not yet been characterized, however, the sequence homology to similar pheromone based plasmids indicated these are involved in mating channel formation (Chen *et al.*, 2008). The PrgE protein has 37% sequence homology to an single stranded binding (ssb) protein identified in a lactococcal phage (Hirt *et al.*, 2005).

Proteins PcfD, PcfE and PcfF belong to the functional class of DNA processing proteins (Figure 7.8 K-M). Interestingly, the corresponding genes *pcfD*, *pcfE* and *pcfF* are located adjacent to each other on the pCF10 plasmid. While, the expression of PcfD increased 2-4 fold within 30-60 minutes post induction (Figure 7.8 K), expression of PcfE increased nearly 4-8 fold in the same time frame (Figure 7.8 L). Surprisingly, PcfF did not show significant activity during pheromone induction with nearly 1.5 fold increase by 2 hours post-induction (Figure 7.8 M), however, PcfF level was abundant enough to be detected by mass-spectrometry. This indicates the *pcfF* expression is constitutive and unresponsive to induction. A drastic drop in dynamic response of protein expression between PcfE and PcfF points towards gene regulation operating in the locus. Even though the PcfA, PcfB and PcfD proteins were identified by mass-spectrometry, the coupling protein PcfC, with corresponding gene *pcfC* located between genes *pcfB* and *pcfD* was not identified.

Three other proteins identified on pCF10, namely, PcfQ, PrgP and PrgZ are encoded from genes present upstream of *prgQ* gene. As expected, these three proteins did not demonstrate dynamic response to induction consistent with lack of transcript activity

as reported in (Hirt *et al.*, 2005). The fact that these were identified by mass-spectrometry indicates that these proteins are expressed at constitutive levels. PrgZ protein is involved in the import of mature iCF10 and cCF10 peptides (Leonard *et al.*, 1996). The functional role of PcfQ is unknown, but shows 100% sequence homology to EFB0041 gene of plasmid pTEF2 (Hirt *et al.*, 2005).

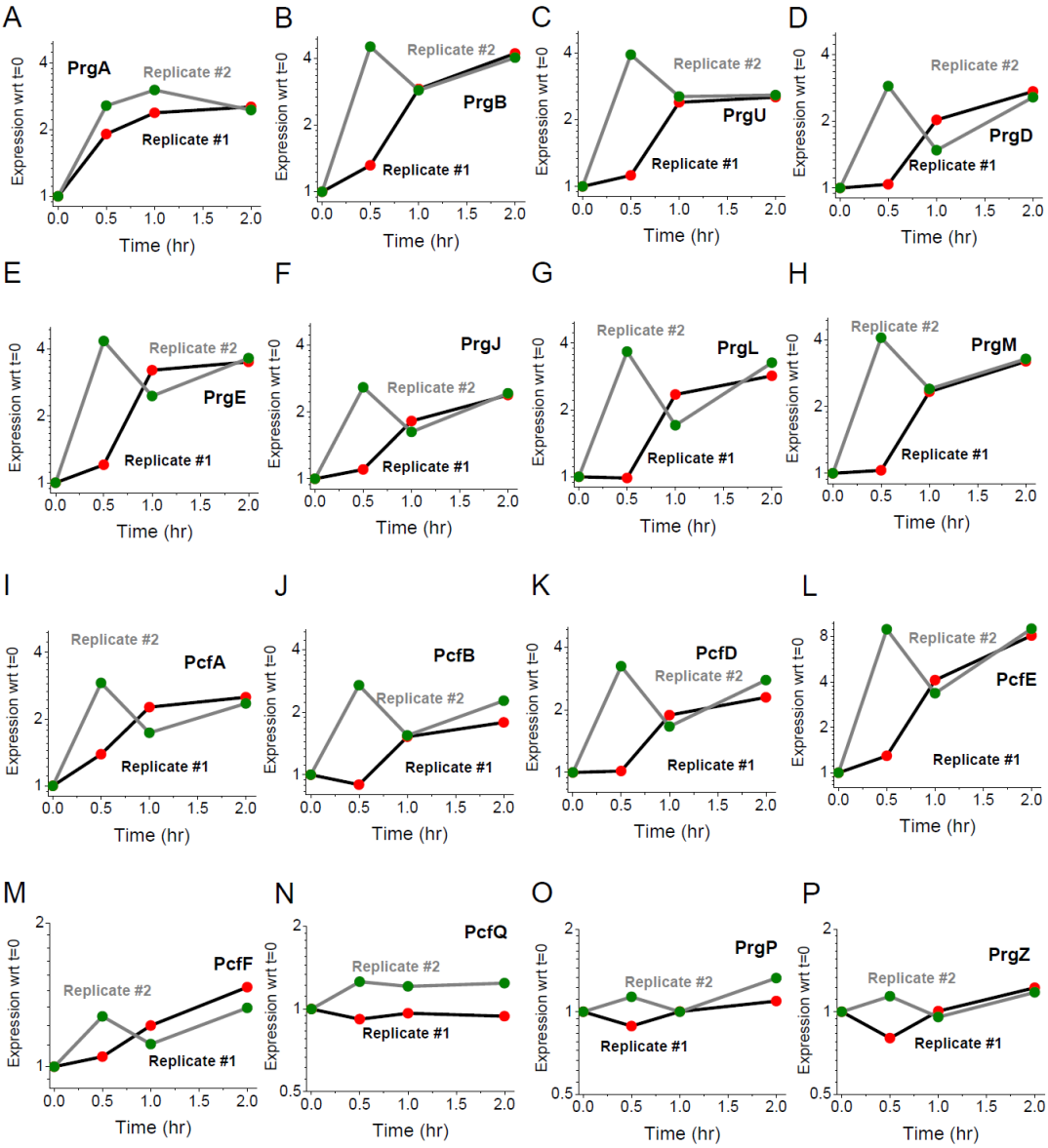


Figure 7.8: Dynamic expression profile of pCF10 based proteins in response to induction. (A-M) Expression profiles of proteins downstream of *prgQ* gene in the order these appear in the pCF10 plasmid. Proteins that were identified include (A) PrgA, (B) PrgB, (C) PrgU, (D) PrgD (E) PrgE, (F) PrgJ, (G) PrgL, (H) PcfA, (I) PcfB, (J) PcfD, (K) PcfE, (L) PrgM, (M) PcfF, **(N-P)** Expression profiles of proteins upstream of *prgQ* gene in plasmid pCF10. (N) PcfQ, (O) PrgP and (P) PrgZ.

The proteins PrgD, PrgE, PrgJ, PrgM, PrgL, PcfA and PcfB encode for proteins that form the mating channel between the donor and the recipient cell (Chen *et al.*, 2008). Again, expression increased significantly (2-4 fold) within 30-60 minutes post-induction for all these proteins (Figure 7.8 D-J). Most of these proteins have not yet been characterized, however, the sequence homology to similar pheromone based plasmids is summarized in Table 2.1. The PrgE protein has 37% sequence homology to an single stranded binding (ssb) protein identified in a lactococcal phage (Hirt *et al.*, 2005).

Proteins PcfD, PcfE and PcfF belong to the functional class of DNA processing proteins (Figure 7.8 K-M). Interestingly, the corresponding genes *pcfD, pcfE* and *pcfF* are located adjacent to each other on the pCF10 plasmid. While, the expression of PcfD increased 2-4 fold within 30-60 minutes post induction (Figure 7.8 K), expression of PcfE increased nearly 4-8 fold in the same time frame (Figure 7.8 L). Surprisingly, PcfF did not show significant activity during pheromone induction with nearly 1.5 fold increase by 2 hours post-induction (Figure 7.8 M), however, PcfF level was abundant enough to be detected by mass-spectrometry. This indicates the *pcfF* expression is constitutive and unresponsive to induction. A drastic drop in dynamic response of protein expression between PcfE and PcfF points towards gene regulation operating in the locus. Even though the PcfA, PcfB and PcfD proteins were identified by mass-spectrometry, the coupling protein PcfC, with corresponding gene *pcfC* located between genes *pcfB* and *pcfD* was not identified.

Three other proteins identified on pCF10, namely, PcfQ, PrgP and PrgZ are encoded from genes present upstream of *prgQ* gene. As expected, these three proteins did not demonstrate dynamic response to induction. The fact that these were identified by mass-spec indicates that these proteins are expressed at constitute levels. PrgZ protein is involved in the import of mature iCF10 and cCF10 peptides (Leonard *et al.*, 1996). The functional role of PcfQ is unknown, but shows 100% sequence homology to EFB0041 gene of plasmid pTEF2 (Hirt *et al.*, 2005).

7.3.9 Dynamics of protein expression from chromosome in response to induction

The expression in response to induction of nearly 381 proteins encoded by chromosomal genes were identified by mass-spectrometry. During the experiment, cells were maintained in the exponential phase (Figure 7.5 A). As expected, majority of the proteins identified (~82.2 %) were not differentially expressed in response to induction. However, approximately 17.8 % of the chromosomal proteins were differentially expressed in both biological replicates, with nearly 4.4 % of the proteins up-regulated and 13.4 % down-regulated in response to induction. A cut-off of 1.5 fold change was considered for differential expression (Jayapal *et al.*, 2008; Jayapal *et al.*, 2010). The up-regulated and down-regulated proteins are listed in Table 7.3 and 7.4 respectively and dynamics of protein expression post-induction of a representative set in shown in Figure 7.9.

A set of ABC transporters involved in peptide transport and peptide-secretion related proteins were found to be up-regulated, implying the cell responds to inducer by producing more transporters. Most of the up-regulated proteins are related to central energy metabolism, including, carbohydrate metabolism (phosphoglycerate mutase 1, bifuncational acetaldehyde-CoA/alcohol dehydrogenase, fumarate reducatase flavoprotein subunit and glucosamine-fructose-6-phosphate aminotransferase), energy metabolism such as oxidative phosphorylation (NADH oxidase), nucleotide metabolism (ionisine 5'-monophosphate dehydrogenase), nitrogen metabolism (glutamate dehydrogenase) and lipid/fatty acid metabolism ((3R)-hydroxymyristoyl-ACP dehydratase). A pheromone binding protein (NCBI-gi: 29377499) and TraG family protein (NCBI-gi: 29377919) were found to be especially up-regulated between 2-4 fold in biological replicates (Figure 7.9 G and H).

On the other hand, few proteins related to metabolism were also found to be down-regulated, especially those corresponding to glycerol metabolism (glycerol kinase, alpha-glycophosphate oxidase, glycerol-uptake facilitator and diacylglycerol kinase catalytic subunit). Contrary to expectation, few protein corresponding to stress response were found to be down-regulated, such as, heat-shock proteins DnaK and GrpE, cold-shock domain containing protein (NCBI-gi: 29377389), 50s ribosomal protein L25/ general stress protein Ctc (NCBI-gi: 29375409) and Chaperonin (NCBI-gi: 29377119). The cell-division protein DivIVA was found to be down-regulated. Unlike few

pheromone binding proteins and transport proteins that were found to be up-regulated, a pheromone pre-cursor lipoprotein cAM 373 was found to down-regulated.

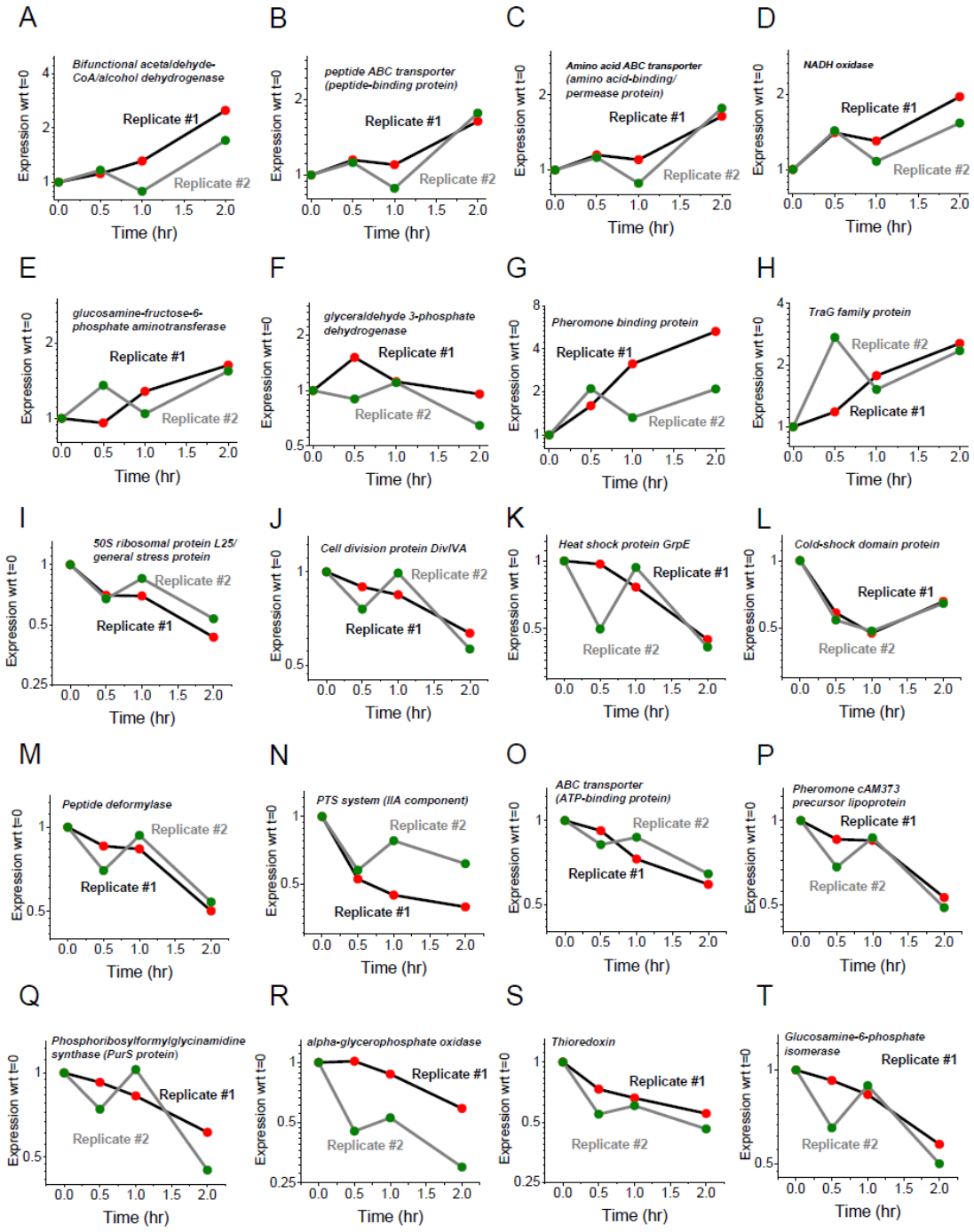


Figure 7.9 Dynamic expression profile of chromosome based proteins in response to induction. Expression profiles of representative **(A-H)** up-regulated proteins and **(I-T)** down-regulated proteins.

Table 7.3: List of up-regulated proteins expressed from the chromosome in response to induction

S. No.	NCBI Gene ID	Gene name (*Enterococcus faecalis* V583)	Pathway description (KEGG)		
			Level1	Level2	Level 2
1	gi: 29374841	phosphoglycerate mutase 1	Metabolism	Carbohydrate Metabolism	Glycolysis/ Gluconeogenesis
2	gi: 29375354	amino acid ABC transporter, amino acid-binding/ permease protein	-	-	-
3	gi: 29375484	bifunctional acetaldehyde-CoA/alcohol dehydrogenase	Metabolism	Carbohydrate Metabolism	Glycolysis/ Gluconeogenesis
4	gi: 29375982	glutamate dehydrogenase	Metabolism	Amino Acid Metabolism	Alanine, aspartate and glutamate metabolism
5	gi: 29376091	glyceraldehyde 3-phosphate dehydrogenase	Metabolism	Carbohydrate Metabolism	Glycolysis/Gluconeogenesis
6	gi: 29376148	NADH oxidase	Metabolism	Energy Metabolism	Oxidative phosphorylation
7	gi: 29376472	hypothetical protein EF1947	-	-	-
8	gi: 29376660	glucosamine--fructose-6-phosphate aminotransferase	Metabolism	Amino Acid Metabolism	Alanine, aspartate and glutamate metabolism
9	gi: 29377044	fumarate reductase flavoprotein subunit	Metabolism	Carbohydrate Metabolism	Citrate cycle (TCA cycle)
10	gi: 29377086	acyl carrier protein, putative	-	-	-
11	gi: 29377214	thioredoxin reductase/glutathione-related protein	Metabolism	Nucleotide Metabolism	Pyrimidine metabolism
12	gi: 29377215	alkyl hydroperoxide reductase, C subunit	-	-	-
13	gi: 29377344	(3R)-hydroxymyristoyl-ACP dehydratase	Metabolism	Lipid Metabolism	Fatty acid biosynthesis
14	gi: 29377499	pheromone binding protein	Environmental Information Processing	Membrane Transport	ABC transporters
15	gi: 29377562	peptide ABC transporter, peptide-binding protein	Environmental Information Processing	Membrane Transport	ABC transporters
16	gi: 29377734	inosine 5'-monophosphate dehydrogenase	Metabolism	Nucleotide Metabolism	Purine metabolism
17	gi: 29377919	TraG family protein	Environmental Information Processing	Signal Transduction	Bacterial secretion system

Table 7.4: List of down-regulated proteins expressed from the chromosome in response to induction

S. No.	NCBI Gene ID	Gene name (*Enterococcus faecalis* V583)	Pathway description (KEGG)		
			Level 1	Level 2	Level 3
1	gi: 29374729	short chain dehydrogenase/reductase family oxidoreductase	-	-	-
2	gi: 29374730	hypothetical protein EF0077	-	-	-
3	gi: 29374732	gls24 protein	-	-	-
4	gi: 29374733	gls24 protein	-	-	-
5	gi: 29374756	ornithine carbamoyltransferase	Metabolism	Amino Acid Metabolism	Arginine and proline metabolism
6	gi: 29374824	deoxyribose-phosphate aldolase	Metabolism	Carbohydrate Metabolism	Pentose phosphate pathway
7	gi: 29374827	hypothetical protein EF0177	-	-	-
8	gi: 29374828	ABC transporter, ATP-binding protein	-	-	-
9	gi: 29375084	OsmC/Ohr family protein	-	-	-
10	gi: 29375094	superoxide dismutase, Mn	-	-	-
11	gi: 29375097	glucosamine-6-phosphate isomerase	Metabolism	Carbohydrate Metabolism	Amino sugar and nucleotide sugar metabolism
12	gi: 29375363	hypothetical protein EF0770	-	-	-
13	gi: 29375409	50S ribosomal protein L25/general stress protein Ctc	Genetic Information Processing	Translation	Ribosome
14	gi: 29375584	cell division protein DivIVA	-	-	-
15	gi: 29375727	cell division protein DivIVA, putative	-	-	-
16	gi: 29375755	hypothetical protein EF1180	-	-	-
17	gi: 29375786	acetolactate synthase	Metabolism	Amino Acid Metabolism	Valine, leucine and isoleucine biosynthesis
18	gi: 29375875	heat shock protein GrpE	-	-	-
19	gi: 29375876	dnak protein	Genetic Information Processing	Folding, Sorting and Degradation	RNA degradation
20	gi: 29375907	pheromone cAM373 precursor lipoprotein	-	-	-
21	gi: 29375935	hypothetical protein EF1368	-	-	-
22	gi: 29375972	thioredoxin	-	-	-
23	gi: 29376123	hypothetical protein EF1560	-	-	-

24	gi: 29376175	CoA-binding domain-containing protein	-	-	-
25	gi: 29376268	carbamoyl phosphate synthase large subunit	Metabolism	Nucleotide Metabolism	Pyrimidine metabolism
26	gi: 29376298	HPr kinase/phosphorylase	-	-	-
27	gi: 29376327	phosphoribosylaminoimidazole synthetase	Metabolism	Nucleotide Metabolism	Purine metabolism
28	gi: 29376330	phosphoribosylformylglycinamidine synthetase I	Metabolism	Nucleotide Metabolism	Purine metabolism
29	gi: 29376331	phosphoribosylformylglycinamidine synthase, PurS protein	Metabolism	Nucleotide Metabolism	Purine metabolism
30	gi: 29376340	hypothetical protein EF1794	-	-	-
31	gi: 29376341	putative lipoprotein	-	-	-
32	gi: 29376454	glycerol uptake facilitator protein	-	-	-
33	gi: 29376455	alpha-glycerophosphate oxidase	Metabolism	Lipid Metabolism	Glycerophospholipid metabolism
34	gi: 29376456	glycerol kinase	Metabolism	Lipid Metabolism	Glycerolipid metabolism
35	gi: 29376504	universal stress protein			
36	gi: 29376923	guanosine 5'-monophosphate oxidoreductase	Metabolism	Nucleotide Metabolism	Purine metabolism
37	gi: 29376992	GcvH family protein	-	-	-
38	gi: 29377088	PTS system, IIA component	Metabolism	Carbohydrate Metabolism	Fructose and mannose metabolism
39	gi: 29377092	F0F1 ATP synthase subunit epsilon	Metabolism	Energy Metabolism	Oxidative phosphorylation
40	gi: 29377106	hypothetical protein EF2621	-	-	-
41	gi: 29377107	hypothetical protein EF2622	-	-	-
42	gi: 29377119	chaperonin, 10 kDa	-	-	-
43	gi: 29377128	diacylglycerol kinase catalytic subunit	-	-	-
44	gi: 29377269	hypothetical protein EF2797	-	-	-
45	gi: 29377373	hypothetical protein EF2909	-	-	-
46	gi: 29377389	cold-shock domain-contain protein	-	-	-
47	gi: 29377524	peptide deformylase	-	-	-
48	gi: 29377586	hypothetical protein EF3132	-	-	-
49	gi: 29377628	succinyl-diaminopimelate desuccinylase	Metabolism	Amino Acid Metabolism	Lysine biosynthesis
50	gi: 29377677	Dps family protein	-	-	-
51	gi: 29377744	myosin-cross-reactive antigen	-	-	-

7.4 Discussion

Pheromone-inducible conjugation plays a significant role in the dispersal of antibiotic resistance and virulence determinants by enterococci, and contributes to their frequent involvement in hospital-acquired infections (Dunny *et al.*, 1978; Dunny, 2007). The pCF10 conjugation machinery serves as a model of complex regulatory interactions, including, competing antagonistic activities of two secreted signaling peptides, multi-layered intracellular regulatory circuits both at the levels of transcription initiation and post-transcriptional regulation to control conjugation (Figure 7.1) (Bensing *et al.*, 1997; Chatterjee *et al.*, 2011; Dunny and Johnson, 2011; Johnson *et al.*, 2010; Shokeen *et al.*, 2010). Besides the biological significance of pheromone inducible conjugation of pCF10 for spread of antibiotic resistance among the bacterial population, the pCF10 system serves as a useful paradigm for cell-cell signaling. The method of cell-cell signaling for the pCF10 system is distinct from the conventional quorum sensing defined elsewhere (Kleerebezem *et al.*, 1997; Waters and Bassler, 2005), in that, two signaling peptides are used to sense two different populations. We postulate that while cCF10 serves as an indicator for the recipient population, iCF10 serves as an indicator of donor population. We show that the inhibitor molecule acts a quorum sensing molecule which synchronizes gene expression of donor population to turn-OFF conjugation.

Previous studies have shown multiple regulation mechanisms control the expression from the *prgQ-prgX* switch give rise to a bistable switch to conjugation. Mathematical model predicts that such a bistable response will shift requiring higher

level of inducer at an increased iCF10 quorum (Figure 7.2 A). However, the process of conjugation is a dynamic process with the levels of cCF10 and iCF10 continuously changing both in the intracellular and extracellular environment. Experimental data showed that expression of Q_L, *prgB, pcfC* and *pcfG* RNA exhibit a pulse of expression in response to induction with nearly 2-3 orders increase in expression 30 minutes, followed by a significant decrease in expression within 2 hours post-induction. Contrary to expectation, turning-OFF response was also observed at high levels of induction (50 ng/mL, Figure 7.5 B-E), though, the turning-OFF behavior was more pronounced at induction with lower levels of cCF10 (0.05 ng/mL, Figure 7.5 B-E) in the time frame of the experiment (Figure 7.5). This is indicative of the fact that induction causes production and accumulation of iCF10 in the medium, once a threshold iCF10 concentration is reached it then counteracts the effect of cCF10 and causes down-regulation of conjugation–related transcripts. At low levels of induction, the levels of iCF10 produced is enough to force the system to return to initial OFF state within 2 hours post induction, whereas, at the higher levels of induction, the amount of iCF10 produced may not be sufficient to fully turn-OFF the entire population, however the downward trend of Q_L and other downstream transcripts is clearly apparent (Figure 7.5 B-E).

This is consistent with model prediction, that at higher iCF10 quorum, expression of Q_L RNA and downstream transcripts of *prgB*, *pcfC* and *pcfG* demonstrate a fall in transcript expression due to increased inhibitory effect of iCF10 (Figure 7.2 B-C and Figure 7.3). Our data is consistent with previously reported transcriptome analysis using a

pCF10 based mini-microarray, where a similar "pulse-like" dynamic expression of downstream pCF10 genes was observed (Hirt *et al.*, 2005). However, in this study, the induction was performed at nearly physiological levels. In the current study we show that such a pulse like response of pCF10 transcripts is maintained even at very high levels of induction (Figure 7.5 B-E). The fact that such a pulse-like response of transcript expression post-induction is lost in absence of endogenous iCF10 production in the ΔEep mutant (Figure 7.7) and is restored when the culture is supplemented with iCF10, clearly highlights the role of iCF10 quorum in controlling the turning OFF response. Previously, it has been shown that plasmid transfer between donor and recipients reaches a peak of transfer activity followed by a steady decline in transfer ability of donors. Such a shut-down of pCF10 transfer activity is consistent with the finding reported here. The increased iCF10 quorum produced by old and newly converted donor cells, may provide an explanation for the fact that not all recipients in a population get converted to donor cells (Hirt *et al.*, 2005).

On one hand, a highly dynamic transcript expression pattern was observed in response to pCF10 induction, on the other hand proteins identified from downstream conjugation-related genes showed a steady ramping-up expression profile in the time course of the experiment, with significant expression between 30 and 60 minutes post-induction, demonstrating a time lag of approximately 30 minutes behind transcript expression. A steady protein expression by 2 hours when transcript expression begins to subside indicates that the pCF10 proteins identified tend to last longer in the donor cell

than the time course of conjugation. Though the proteomic study was carried out at high cCF10 induction (50 ng/mL), similar stable PrgB expression was observed when induction was carried out moderate levels of cCF10 (Hirt *et al.*, 2005). Interestingly, except for PcfF, the accessory factor which functions in conjunction with PcfG to cleave pCF10 DNA at the *oriT* sequence (Chen *et al.*, 2008), other proteins that are encoded by genes present downstream of *prgQ* gene showed similar dynamic response to induction, whereas, proteins encoded by gene upstream of *prgQ* gene lacked dynamic response to induction consistent with transcript data reported in (Hirt *et al.*, 2005) (Figure 7.8). This is consistent with the notion that induction affects the response of *prgQ-prgX* genetic switch, causing expression of Q_L RNA, which further induces expression of the downstream genes via post-transcriptional mechanism (Bensing *et al.*, 1997). Interestingly, while significant expression of *pcfC* and *pcfG* transcript expression was observed at 50 ng/mL, the corresponding PcfC and PcfG proteins were not identified using mass-spectrometry. This could indicate that these proteins are less abundant than those that were identified in this study. This hints towards the possibility of a rate limiting effect of weakly expressed proteins in assembly of the conjugation machinery.

Bacteria communicate with one another by via Quorum-sensing, which involves the production and release of chemical signal molecules, whose external concentration increases as a function of increasing cell-population density (Kleerebezem *et al.*, 1997; Miller and Bassler, 2001; Waters and Bassler, 2005). Accumulation of signal concentration above a minimal threshold, results in its detection by bacteria with

consequent effect on altering gene expression and coordinate cell behavior on a population-wide scale. Here, we described the role of iCF10, the inhibitor signaling peptide for pCF10 based system, as quorum-sensing signal with the aim of turning-OFF conjugation at a population-wide level.

In the present study we showed that induced pCF10 plasmid-bearing cells express genes downstream of *prgQ* gene encoding various proteins involved in transfer of the plasmid. Induction also results in elevated expression of *prgQ* transcripts which encodes the secreted iCF10. The accumulation of iCF10 in the extracellular environment due to concerted action of neighboring donor cells amplifies the iCF10 concentration, thus giving rise to negative feedback loop and consequently allowing the system to quickly turn-OFF conjugation. Quorum sensing has been well-characterized in gram-negative bacteria, such role of N-acyl homoserine lactones (AHL) in coordinated expression of bioluminescence genes in *Vibrio fischeri* . Similarly, several processes in gram-positive bacteria are known to be regulated in a cell-density depended manner, including development of genetic competence in *Bacillus subtilis* (Dubnau, 1991), antibiotic production in *Streptomyces gresius* (Chater and Horinouchi, 2003), virulence in *Staphylococcus aeurus* (Morfeldt *et al.*, 1995; Novick *et al.*, 1993). In most of these quorum-sensing systems, the role of the chemical signal typically involves turning-ON of the cognate genetic switch. In pCF10 system, the role of quorum-sensing signal is reversed, in that, iCF10 is used for turning-OFF the *prgQ-prgX* switch controlling conjugation. Similar negative role of has been shown for Farnesol, a quorum sensing

signal that has been shown to inhibit biofilm formation in *Candida albicans* (Ramage *et al.*, 2002). Such a quorum-sensing based mechanism by which donor cells turn OFF conjugation is critical for both efficient transfer of plasmid and for minimizing the metabolic burden of the plasmid transfer process. The present study changes our view of iCF10 signaling molecule as a quorum-sensing molecule. Use of such inhibitors of conjugation show promise as compounds for drug development (Lujan *et al.*, 2007).

7.5 Materials and Methods

7.5.1 Plasmids, strains, and growth conditions: OG1Sp+pCF10 cells were incubated in complete M9 medium containing 10 μg/ml Tetracycline and 100 μg/ml Spectinomycin. Overnight cultures were diluted 1:10 and split into 5 tubes containing five different concentrations of cCF10, i.e. 0, 0.05, 0.5, 5 and 50 ng/mL. Samples of cells (~600 μL) were collected at time points 0, 15, 30, 60 and 120 minutes. RNA was extracted using the RNeasy kit (Qiagen), followed by DNAse treatment using Turbo free kit (Life Science Technologies). cDNA was synthesized using SIII First Strand Synthesis system (Life Science Technologies). Random hexamers were used for the reverse transcription. QPCR was performed using Brilliant SYBR Green mix (Agilent) on MXPro 3000 machine. Dynamics of expression of 6 genes were probed, namely: *prgX*, Q_L, *prgB, pcfC, pcfG* and *Gyrase* (housekeeping gene). The following data represent one biological replicate.

7.5.2 RNA extraction and QPCR: 0.6 mL of culture was added to 1.2 mL of Bacteria RNA protect reagent (Qiagen), followed by 5 min incubation. Samples were flash frozen in Ethanol and dry ice and stored at -80°C. Cells were lysed with lysozyme (30 mg/mL)

in TE buffer + Mutanolysin (500 U/mL). RNA was prepared using RNeasy kit (Qiagen). 5 μg of RNA was subjected to rigorous DNAse treatment with Turbo DNAse (Ambion) as per manufacturer's instructions, followed by reverse transcription of 100 ng of DNA free RNA using Superscript III first strand synthesis kit (Invitrogen 18080-051) with random primers. Approximately 5 ng of cDNA was used for quantitative PCR using Brilliant SYBR Green QPCR Master Mix (Stratagene, Agilent Technologies) using gene specific primers (SI text) on real-time MX3000P instrument (Stratagene, La Jolla, CA). Transcript levels were determined using the ΔCt method with respect to *GyraseB*, a moderately expressed housekeeping gene used as a reference.

7.5.3 iTRAQ and Protein Extraction: OG1Sp+pCF10 cells were incubated in complete M9 medium containing 10 μg/mL Tetracycline and 100 μg/mL Spectinomycin. Overnight cultures were diluted 1:10 and was split into 4 tubes containing pre-warmed complete M9 medium and allowed to grow at 37°C for 30 minutes, followed by addition of 50 ng/mL of cCF10 to all tubes. Samples of cells (~10 mL) were collected at time points 0, 30, 60 and 120 minutes. The cells were pelleted and flash frozen in dry ice and 95% Ethanol bath. Samples were stored in -80°C.The cell pellets were washed twice in 700 μL of cracking buffer (50 mM MOPS, pH 7.5, 5% v/v protease inhibitor (Sigma Aldrich)) and centrifuged at 6000×g at 4°C for 5 minutes. The cell pellets were re-suspended in 700 μL of cracking buffer and subjected to sonication for 30 seconds × (5 times) with 1 minute time interval on ice to enable Lysozyme-free cell lysis.

The samples were centrifuged at 13,000 rpm to collect supernatant. Protein concentration was quantified using BCA Assay (Thermo Scientific). Approximately 100-200 μg of protein sample was precipitated overnight in 100% Ethanol at -20°C. Protein was pelleted by centrifugation at 6000×g at 4°C for 10 minutes. The precipitated protein pellet was re-suspended in 25 μL of TEAB dissolution buffer (Applied Biosystems) and quantified using BCA Assay (Thermo Scientific). A final amount of 40 μg of protein in 29 μL of TEAB dissolution buffer was submitted for proteomic analysis. The same procedure was followed for two biological replicates. Identification and relative quantitation of proteins in response to induction of pCF10 carrying donor cells was performed using 8-plex iTRAQ reagents (Applied Biosystems) in conjunction with liquid chromatography (LC) and mass spectrometry (MS). A total of 8 samples of extracted proteins from two biological replicates (4 time points each) were subjected to buffer exchange using a YM3 Microcon (Millipore), normalized for concentration of protein, reduced, alkylated, trypsin digested independently in parallel and labeled with iTRAQ reagents 113, 114, 115, 116, 117, 118 and 121 respectively. The labeled peptides from both samples were pooled and subjected to 2D LC-MALDIMS/MS using Tempo™ LC MALDI spotting system, and 4800 MALDI TOF/TOF™ analyzer at the Center for Mass Spectrometry at the University of Minnesota.

7.5.4 Proteomic Data Analysis: Protein identification and quantitation was carried out using ProteinPilot™ 4.0 software (Applied Biosystems). This software compares the fragmentation pattern of the spectra obtained by MS/MS to the theoretical fragmentation

pattern of proteins in a protein database to identify the proteins. *Enterococcus faecalis* V583 chromosome (NC_004668.1) and pCF10 sequence (AY855841.2) were used to generate the protein database. For relative quantification, ProteinPilot™ was used to calculate the iTRAQ ratio for each reagent pair, and determine an average iTRAQ ratio for each protein. This ratio was used to identify proteins that are up or down regulated in response to induction with cCF10. Data analysis was conducted by exporting all data sets from Protein Pilot to Microsoft Excel (Office 2007). To be classified as significantly altered in expression, the following criteria is considered: (1) the protein had to demonstrate a database match with a enterococcal protein; (2) the protein had to be represented by at least two peptides of differing amino acid sequence at the 95% confidence level, (3) the test to determine if the ratio of protein expression equals 1 had to have a p-value less than 0.05, (4) the protein needed to have an error factor as calculated by Protein Pilot of less than 2, (5) the fold change needed to be greater than or equal to 1.5 or less than or equal to 0.67. (6) The protein needed to match the criteria above for the two separate runs of iTRAQ that was performed on these samples for confirmation of the data.

7.5.5 Mathematical modeling: A set of ordinary differential equations based on pCF10 genetic network were developed (Table 7.1). Steady state and dynamic simulations were performed in MATLAB (version 2008a MathWorks, Natick, MA). Steady state solutions of non-linear algebraic equations were obtained using MATLAB solve function. ODEs were integrated using function *ode23s* in MATLAB to obtain dynamic behavior.

CHAPTER 8

CONVERGENT TRANSCRIPTION IN THE BUTYROLACTONE REGULON IN STREPTOMYCES COELICOLOR CONFERS A BISTABLE GENETIC SWITCH FOR ANTIBIOTIC *BIOSYNTHESIS*

8.1 Summary

cis-encoded antisense RNA's (*cis* asRNA) have been reported to participate in gene expression regulation in both eukaryotic and prokaryotic organisms. Its presence in *Streptomyces coelicolor* has also been reported recently; however, its role has yet to be fully investigated. Using mathematical modeling we explore the role of *cis* asRNA produced as a result of convergent transcription in *scbA-scbR* genetic locus. *scbA* and *scbR* gene pair, encoding repressor–amplifier proteins respectively, mediates the synthesis of a signaling molecule, the γ-butyrolactone SCB1 and controls the onset of antibiotic production. Our model considers that transcriptional interference caused by convergent transcription of two opposing RNA polymerases results in fatal collision and transcriptional termination, which suppresses transcription efficiency and gives rise to truncated transcripts. Additionally, convergent transcription causes sense and antisense

interactions between complementary sequences from opposing strands, rendering the full length transcript inaccessible for translation. We evaluated the role of transcriptional suppression and the antisense effect conferred by convergent transcription on the behavior of *scbA-scbR* system. Stability analysis showed that while convergent transcription affects the system, it is asRNA that confers *scbA-scbR* system the characteristics of a bistable switch in response to the signaling molecule SCB1. With its critical role of regulating the onset of antibiotic synthesis the bistable behavior offers this two gene system the needed robustness to be a genetic switch. The convergent two gene system with potential of transcriptional interference is a frequent feature in various genomes. The possibility of asRNA regulation in other such gene-pairs is yet to be examined.

8.2 Introduction

Transcription from a pair of promoters arranged in face-to-face orientation is ubiquitous both in bacteria and eukaryotes. It leads to a complete or partial overlap between convergent or divergent transcripts. Widespread convergent transcription leads to a large number of *cis* asRNA in eukaryotic genomes including human (Yelin *et al.*, 2003), mouse (Katayama *et al.*, 2005), *Drosophila* (Misra *et al.*, 2002), *A. thaliana* (Wang *et al.*, 2005), and yeast (David *et al.*, 2006). Many of these *cis* asRNA's are non-coding, but have been shown to participate in regulation (Georg *et al.*, 2009; Katayama *et*

al., 2005; Prasanth and Spector, 2007). Recent genomic analysis in bacteria has revealed a plethora of *cis*-encoded non-coding RNA in many species, including *E. coli* (Dornenburg *et al.*), *B. subtilis* (Rasmussen *et al.*, 2009), and *Mycobacterium tuberculosis* (Arnvig and Young, 2009). The discovery of *cis* asRNA in the model streptomycete *Streptomyces coelicolor* was reported only recently (D'Alia *et al.*; Panek *et al.*, 2008; Swiercz *et al.*, 2008). Despite the increasing evidence of antisense transcription, the regulatory role of convergent transcription has been not been investigated in *S. coelicolor*. In this work, the role of convergent transcription in the *scbA-scbR* system is evaluated.

The soil dwelling organism, *S. coelicolor* uses its arsenal of antibiotics to compete with other organisms in the environment. Its production of antibiotics is regulated by the synthesis of γ-butyrolactones, members of the quorum sensing-type family of signaling molecules (Hughes and Sperandio, 2008; Takano, 2006), which are found in many *Streptomyces* species, including *Streptomyces virginiae* (Kinoshita *et al.*, 1997), *Streptomcyes lavendulae* (Kitani *et al.*, 2001; Ruengjitchatchawalya *et al.*, 1995), and *Streptomyces clavuligerus* (Kim *et al.*, 2004). In *S. coelicolor* A3(2), three kinds of γ-butyrolactones have been identified which serve to synchronize the onset of antibiotic synthesis within the population (Horinouchi, 2002; Hsiao *et al.*, 2009; Takano *et al.*, 2000), among these *S. coelicolor* butanolide 1 (SCB1) is abundantly found. These SCBs regulate antibiotic biosynthetic clusters controlling synthesis of blue pigmented

actinorhodin (Act) (Takano *et al.*, 2001), red pigmented undecylprodigiosin (Red) (Takano *et al.*, 2001) and yellow pigment yPCK (Gottelt *et al.*).

The onset of antibiotic production has to be a tightly regulated process as antibiotics can be lethal even to their producers. The genetic switch controlling the transition from a non-producing state to an antibiotic producing state must be robust. Previously, it has been shown that *S. coelicolor* changes from a vegetative growth state without antibiotic production (OFF state) to an antibiotic producing state (ON state), upon induction with SCB1. This leads to amplification of the γ-butyrolactone SCB1 signal, resulting in a switch-like transition (Mehra *et al.*, 2008c; Takano *et al.*, 2001). A gene pair, *scbA* (SCO6266) and *scbR* (SCO6265), convergently transcribed from a set of face to face promoters, pA and pR respectively, regulates the biosynthesis of SCB1 (Figure 8.1) (Takano *et al.*, 2001).

The s*cbR* gene encodes for a cytoplasmic receptor protein, ScbR (R), which has a γ-butyrolactone binding domain at its C-terminal and a DNA binding domain in its N-terminal (Natsume *et al.*, 2004; Onaka *et al.*, 1997). In absence of SCB1, ScbR auto-represses itself and represses *scbA* through binding at the O_R and O_A operator sites (Figure 8.1 A-B), flanking promoter pR and pA respectively (Takano *et al.*, 2001; Takano, 2006). Additionally, in absence of signaling molecules, ScbR represses expression from the cryptic type I polyketide synthase gene cluster (*cpk*) by directly binding to promoter region of *cpkO*, the activator of the *cpk* gene cluster (Takano *et al.*,

2005). The *scbA* gene encodes for ScbA, a AfSA homologue, a key enzyme in synthesis of A-factor, the γ-butyrolactones in *S.gresius* (Kato *et al.*, 2007). ScbA has homology to fatty acid synthases and has been shown to be involved in the production of SCB1 from glycerol derivatives and β-keto acid derivatives as precursors (Hsiao *et al.*, 2007). At high concentrations, SCB1 binds to ScbR to form a SCB1-ScbR (CR) complex, thereby relieving its self-repression (Takano *et al.*, 2001).

The proposed regulatory mechanism of the system was previously analyzed using a mathematical model and was shown to exhibit a bistable response of regulatory repressor ScbR levels at varying concentrations of the signaling molecule SCB1 (Mehra *et al.*, 2008). In the model, ScbA and ScbR were postulated to form a protein complex (ScbA-ScbR) which acts as a positive regulator of ScbA (Mehra *et al.*, 2008), thus upregulating SCB1 synthesis. ScbA-ScbR formation was a key component contributing to the bistability. However, lack of experimental evidence for the ScbA-ScbR complex (Mehra *et al.*, 2008) prompted us to look for alternative mechanisms that could confer bistability to this system.

The *scbA* and *scbR* genes overlap by 53 bp from their respective transcription start sites (Takano *et al.*, 2001) (Figure 8.1 A), resulting in a possible head-to-head collision of converging RNA polymerases (RNAPs) either between both elongating RNAP or between an elongating RNAP and RNAP stationed at the opposing promoter serving as a sitting duck for collision. Such a suppressive influence of transcriptional activity of nearby or overlapping genes *in cis* is referred to as transcriptional interference

(TI) (Callen *et al.*, 2004; Palmer *et al.*, 2009; Shearwin *et al.*, 2005; Ward and Murray, 1979). TI caused by RNAP collision leads to transcriptional termination, which results in decrease in expression of full-length RNA from promoters pA and pR and generation of truncated RNA (Figure 8.1 B). In addition to TI, the convergent transcription also generates transcripts that have a segment of complementary sequence, which may incur antisense interactions between sense-antisense full-length *scbR* (*r*) and *scbA* (*a*) transcripts resulting in translational inhibition or mRNA degradation of hybrid RNA complexes (Brantl, 2002) (Figure 8.1 B).

Here we show that convergent transcription from the *scbA-scbR* locus alone, without positive feedback from a hypothetical ScbA-ScbR protein complex, yields a robust bistable genetic switch in response to the signaling molecule SCB1. Similar switches could potentially operate in other two-gene systems arranged in convergent orientation in *S. coelicolor*.

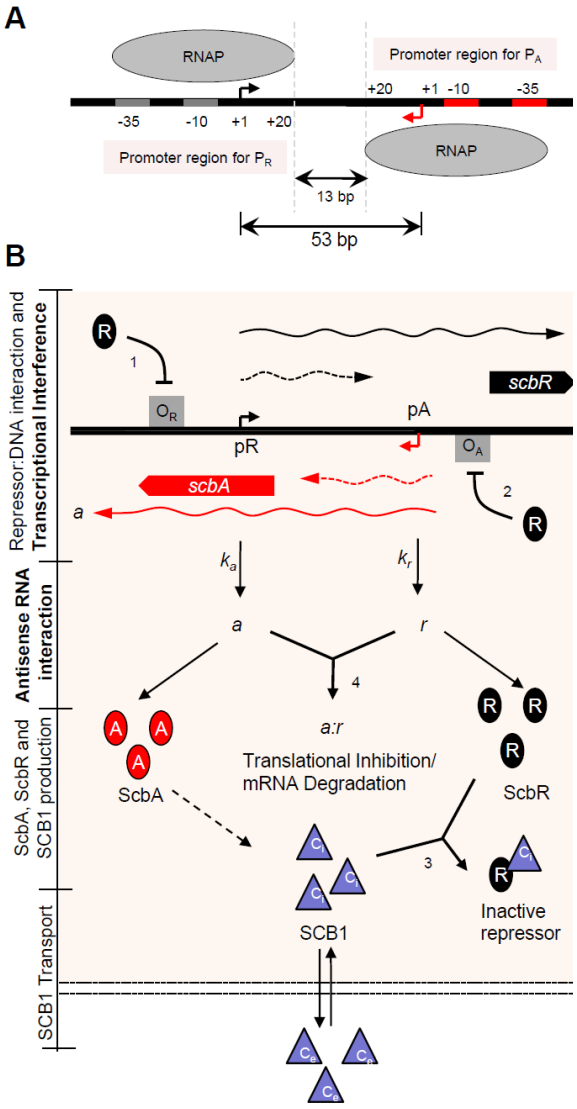


Figure 8.1: Convergent transcription in the *scbA-scbR* gene regulatory network. (A) Schematic of the promoter regions for pA and pR, and footprint of an RNAP at the promoter projected within the overlapping DNA. **(B)** The *scbA-scbR* gene regulatory network. Convergent promoter pA-pR drive expression of genes *scbA* (shown in red) and *scbR* (shown in black) present on the opposite DNA strands (shown by black lines) to give rise to full-length transcripts *a* and *r* (RNA denoted by curved lines) respectively. **Transcriptional Interference model:** Collision between elongating and stationary RNAP within the 53 bp overlapping DNA results in transcriptional termination resulting in the generation of short truncated transcripts (denoted by dashed-curved lines) from promoters pA and pR respectively. Full-length transcripts *a* and *r* result when elongating RNAPs escape collision. **Antisense Regulation:** Hybrid RNA complexes formed between full-length *a* and *r* RNA result in translational inhibition or mRNA degradation. Protein ScbR (R) can repress transcription from pA and pR (indicated by blunt arrows) by binding to operator sites O_A and O_R adjacent to promoters pA and pR respectively. Only full-length transcripts *a* and *r* are translated to protein ScbA (A) and ScbR (R). The intracellular γ-butyrolactone SCB1 (denoted by C_i) is produced from glycerol derivatives and β-keto acid by the enzymatic action of γ-butyrolactone synthase ScbA. SCB1 forms complex with ScbR (C_iR) to sequester its repressive effect. SCB1 diffuses out of the cell in to the extracellular environment and vice versa (denoted by C_e). The reactions are numbered according to equations in Table 8.1.

8.3 Results

8.3.1 Transcriptional interference in scbA-scbR locus

The convergent transcription from pA and pR and the overlapping region of *scbA-scbR* is shown in Figure 8.1 A. The success or failure of each transcription initiation hinges on whether RNAP fires from or binds at promoter pA during the time taken by RNAP from pR to traverse the overlapping DNA and vice versa (Wang *et al.*, 1998). Our model incorporates three mechanisms of transcriptional interference, namely, (i) promoter occlusion, in which a passing RNAP originating at pR blocks access to pA and vice versa, (ii) collision between converging elongating RNAP originating from pA and pR and (iii) sitting duck collisions, in which closed promoter complex at pA is removed by collision with a passing RNAP originating at pR and vice versa (Shearwin *et al.*, 2005; Sneppen *et al.*, 2005).

The footprint of an RNAP bound to the promoter is considered to extend between -50 to +20 bp (Chen *et al.*, 2010; Sneppen *et al.*, 2005). In an event both pA and pR promoters are bound by RNAP, each RNAP can travel a maximum of 13 bp before a collision occurs (Figure 8.1 A). Collision between RNAPs regardless whether elongating or bound to promoter is considered to be fatal for both the polymerases (Sneppen *et al.*, 2005). It is assume that the 3' end-most base of a nascent transcript is 20 bp from the locus of the front end of RNAP (RNAP footprint), and this is used to calculate the length of a truncated RNA due to aborted transcription (Sneppen *et al.*, 2005). Binding of RNAP at one promoter is prevented (occlusion) once an elongating RNAP originating

from the other reaches within 20 bp from the start site of the opposing promoter. With conditions stated above the maximum length of the truncated transcript resulting from aborted transcription would be 13 nt (Figure 8.1 A).

The secondary structure of transcript in the overlapping region (Figure 8.2 A) was predicted using Sfold (Figure 8.2 B). The 100 nt pR transcripts comprises of G-rich single-stranded region, indicated as Stem loop I in Figure 8.2 B. This stem loop is complementary to the RBS of 100 nt pA transcript (Takano *et al.*, 2001), indicating likely sense–antisense RNA interaction with *scbA*. Presence of additional single-stranded regions complementary between *a* and *r* RNA, further enhance the potential of sense–antisense interactions between transcripts from this locus (Brantl, 2007). With the truncated RNA having a maximum size of 13 nt from transcription initiation, and the stem loop region capable of sense-antisense interaction being located between 14-17 nt of pR transcript and 38-43 nt of pA transcript, we conclude that the sense-antisense interaction between the transcripts from pA and pR occur only between full-length mRNAs, but not between a truncated RNA and a full-length transcript.

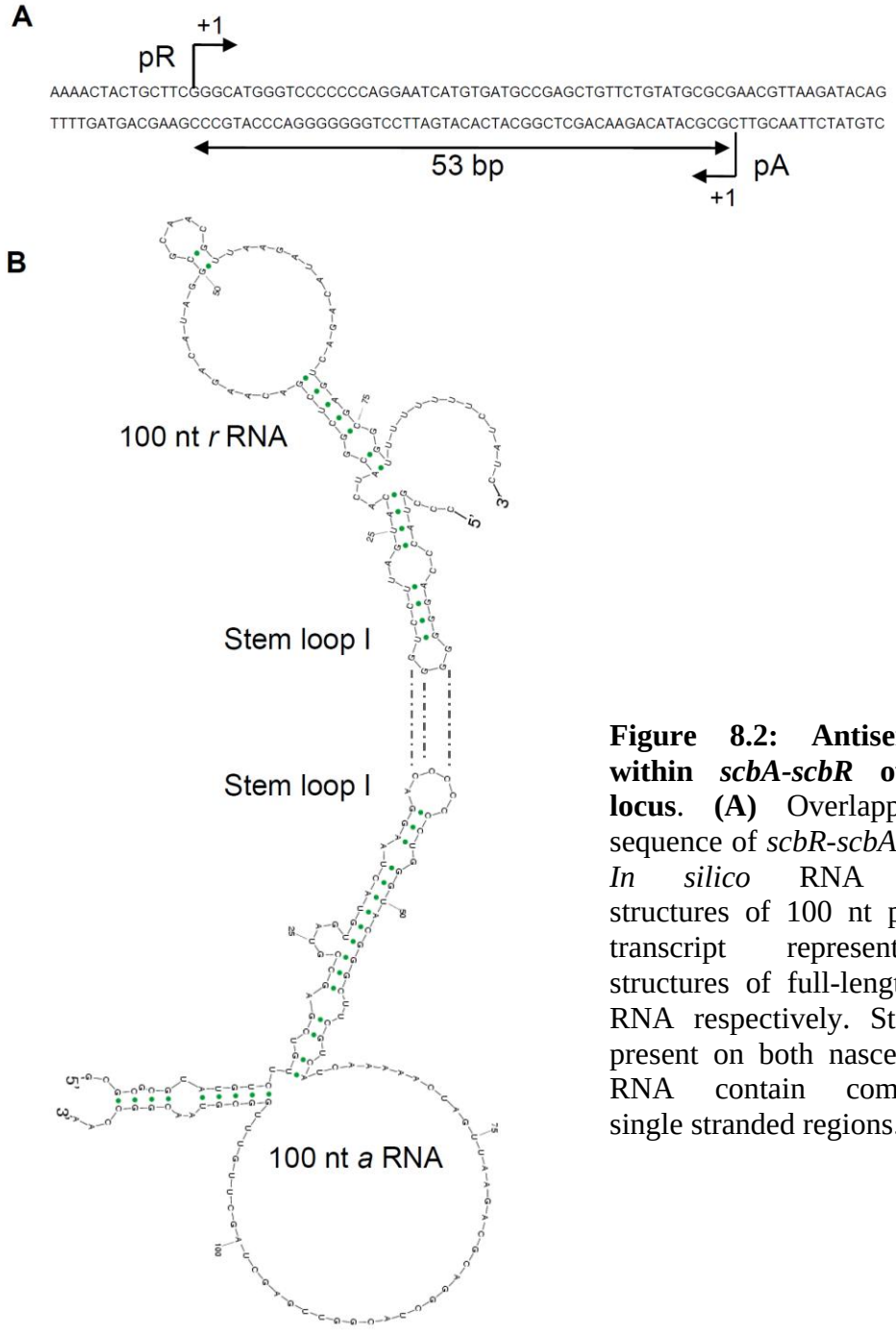


Figure 8.2: Antisense RNA within *scbA-scbR* overlapping locus. **(A)** Overlapping DNA sequence of *scbR-scbA* locus. **(B)** *In silico* RNA secondary structures of 100 nt pR and pA transcript representing the structures of full-length *r* and *a* RNA respectively. Stem loop I present on both nascent *a* and *r* RNA contain complementary single stranded regions.

8.3.2 Mathematical model of convergent transcription in the scbA-scbR locus

A mathematical model for the revised *scbA/scbR* system consisting of a set of 7 differential equations was formulated to describe the model shown in Figure 8.1 B. The binding and hybridization reactions, the equilibrium relationship and the mass action equations are shown in Tables 8.1 and 8.2 (Equations 8.1-8.15). The binding interactions between the ScbR repressor protein (R) and the operators O_R and O_A (Equations 8.1 and 8.2), binding of ScbR to SCB1 (C) (Equation 8.3) are considered. Importantly, the model also considers antisense interaction between full-length transcripts (*a* and *r*) to result in formation of RNA hybrid complexes *a*:*r* (Equations 8.4). Assuming binding of R to operators O_R and O_A reaches rapid equilibrium, the fraction of the unbound operator sites is given by Equations 8.5 and 8.6 respectively (Mehra *et al.*, 2008). The synthesis rate of SCB1 is assumed to be proportional to the ScbA concentration and is given by the first term in Equation 8.13 (Mehra *et al.*, 2008).

In order to estimate the transcription rate from promoters pA and pR, the effect of convergent transcription is taken into consideration. The RNAP binding rate at promoters pA and pR is considered proportional to the concentration of de-repressed operator sites (unbound by R) and is given by Equations 8.7 and 8.8, where $k_{pR-\min}$ and $k_{pA-\min}$ are the basal transcription rates under repressor-bound conditions, and $k_{pR-\max}$ and $k_{pA-\max}$ are the transcription rates under de-repressed (unbound) conditions from promoters pR and

pA, respectively. The RNAP binding rate at pR and pA (respectively denoted by k_{pR} and k_{pA} in Equations 8.7 and 8.8) is then the combined contribution of basal (repressed) and de-repressed rates. This is also the transcription rate of full-length *r* and *a* RNA in the absence of TI effects.

Table 8.1: Reactions in *scbA-scbR* gene network

Equation No.	Equation/Reaction	Description
8.1	$O_R + R \underset{k\_OR}{\overset{k_{OR}}{\rightleftarrows}} O_R\text{-}R$	Reversible binding of ScbR to operator site O_R
8.2	$O_A + R \underset{k\_OA}{\overset{k_{OA}}{\rightleftarrows}} O_A - R$	Reversible binding of ScbR to operator site O_A
8.3	$C_i + R \underset{k\_bCR}{\overset{k_{bCR}}{\rightleftarrows}} CR$	Reversible binding of SCB1 to ScbR
8.4	$a + r \underset{k\_bar}{\overset{k_{bar}}{\rightleftarrows}} a:r$	Reversible binding of full-length transcripts *a* and *r*

Table 8.2 Rates and Mass-action Equations for ScbA-ScbR model

Equation No.	Equation	Description
8.5	$\frac{[O_R]}{[O_{R,T}]} = \frac{K_{OR}}{K_{OR}+[R]}$	Equilibrium relationship between number of unoccupied O_R sites to total number of $O_{R,T}$ sites
8.6	$\frac{[O_A]}{[O_{A,T}]} = \frac{K_{OA}}{K_{OA}+[R]}$	Equilibrium relationship between number of unoccupied O_A sites to total number of $O_{A,T}$ sites
8.7	$k_{pR} = k_{pR-\max}\frac{[O_R]}{[O_{R,T}]} + k_{pR-\min}\left(1-\frac{[O_R]}{[O_{R,T}]}\right)$	RNAP firing rate from promoter pR
8.8	$k_{pA} = k_{pA-\max}\frac{[O_A]}{[O_{A,T}]} + k_{pA-\min}\left(1-\frac{[O_A]}{[O_{A,T}]}\right)$	RNAP firing rate from promoter pA
Mass-action equations for ScbA-ScbR model		
8.9	$\frac{d[r]}{dt} = k_r - k_{bar}[a][r] + k_{\_bar}[a:r] - (k_{dr}+\mu)[r]$	
8.10	$\frac{d[a]}{dt} = k_a - k_{bar}[a][r] + k_{\_bar}[a:r] - (k_{da}+\mu)[a]$	
8.11	$\frac{d[R]}{dt} = k_R[r] - k_{bCR}[C_i][R] + k_{\_bCR}[C_iR] - (k_{dR}+\mu)[R]$	
8.12	$\frac{d[A]}{dt} = k_A[a] - (k_{dA}+\mu)[A]$	
8.13	$\frac{d[C_i]}{dt} = k_C[A] - k_{bCR}[C_i][R] + k_{\_bCR}[C_iR] - k_{se}([C_i]-[C_e]) - (k_{dC}+\mu)[C_i]$	
8.14	$\frac{d[C_iR]}{dt} = k_{bCR}[C_i][R] - (k_{\_bCR}+k_{dCR}+\mu)[C_iR]$	
8.15	$\frac{d[a:r]}{dt} = k_{bar}[a][r] - (k_{\_bar}+k_{dar}+\mu)[a:r]$	

The overall success and failure rate of pR and pA initiated transcription depends on the relative RNAP binding rates, k_{pR} and k_{pA} (Equation 8.7-8.8), the time taken to transition from a closed promoter complex to an elongation complex (τ) and the RNAP

traveling time within the overlapping DNA. To determine the transcription rate of *a*, *r* for use in the ordinary differential equation (ODE) model (Table 8.2), we employed discrete simulation to calculate the formation rate of those RNA species over time. In the discrete simulation, RNAP binds to both pA and pR at time t=0 and starts transcription after a τ=2 second delay (Tang *et al.*, 2009), which is was kept lower than the minimum RNAP binding time interval at promoters pA and pR (2.2 s and 4.8 s respectively) when concentration of R tends to zero. In the simulations, τ for pA and pR was kept the same. In absence of RNAP collision, thereafter, RNAP fires from pR and pA at rates k_{pR} and k_{pA}, respectively. Since both k_{pR} and k_{pA} are a function of R concentration (Equation 8.7-8.8), the simulation was carried out for different R concentrations (Figure 8.3 A-C). The velocity of an elongating RNAP is set at v_0=50 bp/s (Neuman *et al.*, 2003). RNAP is assumed to move along the DNA at a time step of $1/v_0$ for every base. Movement of RNAP was tracked along both strands of DNA within the overlapping region. The model does not consider RNAP pausing for a short stretch of overlapping DNA such as 53 bp. Road blocking effect of repressor ScbR bound to DNA is not considered based on negligible road blocking effect observed in the P_Q-P_X system in *Enterococcus faecalis* (Chatterjee *et al.*, 2011). The parameters used for the simulation are summarized in Table 8.3.

The resulting apparent transcription rates of *a*, *r* to be used in the ODE model (Equation 8.9-8.10) are shown in Figure 8.3 A-B. Without RNAP collision the transcription rate of *r* RNA ($k_r = k_{pR}$) increased about two fold under derepressed

conditions (low R concentration) compared to repressed state (high R concentration). The presence of RNAP collision altered the dynamics of *r* transcription. Instead of having a modest increase at low R concentrations, the transcription rate of *r* (k_r) decreases to nearly zero. The vast majority of RNAP binding to pR are predicted to be knocked off by colliding RNAP from pA because of higher RNAP firing frequency from pA (3-fold higher than pR).

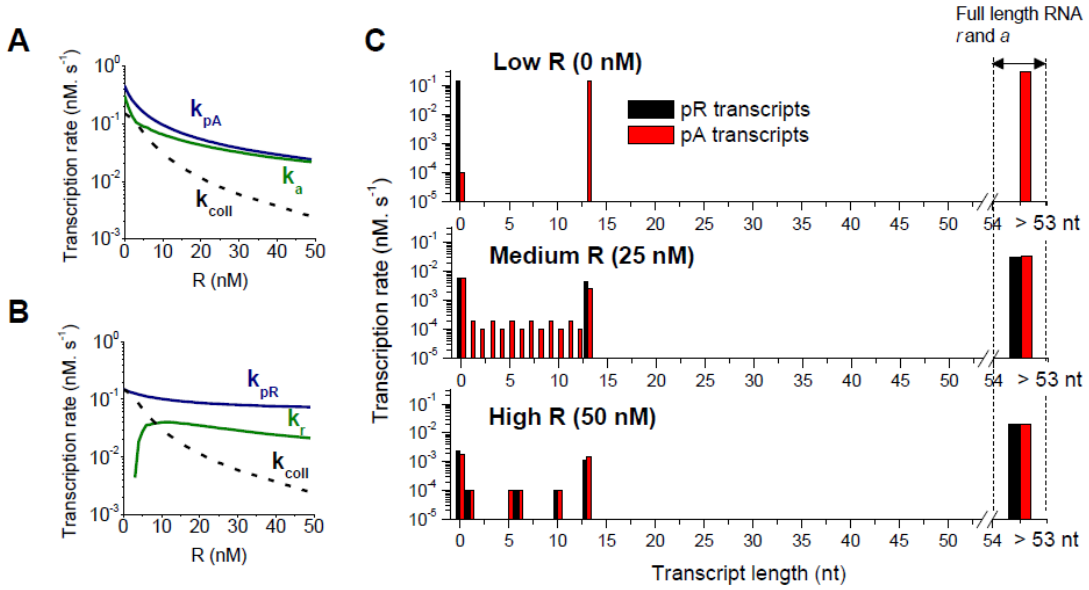


Figure 8.3: Transcriptional Interference within *scbA-scbR* locus. (A-B) Transcription rates k_a and k_r from promoters pA and pR respectively in presence of TI, the RNAP firing rates k_{pA} and k_{pR} from promoters pA and pR respectively in absence of TI, and the net rate of RNAP collisions k_{coll} due to sitting duck collisions and collision between elongating RNAP, shown for different concentrations of repressor ScbR. **(C)** Transcription rate of different sizes of truncated RNA (<53 nt) and full-length RNA (*a* and *r*) from promoters pR and pA for different levels of repressor R.

At such a low level of R, full length *r* is reduced drastically. As expected, the RNAP collision rate increases with the increasing RNAP binding rates when R

concentration decreases (Figure 8.3 A-B), consequently rate of production of very short truncated RNA (<13 nt) increases (Figure 8.3 C). The size distribution of the truncated RNA varies over the range of transcription initiation rates (Figure 8.3 C), however, these short truncated RNA lack secondary structure and thus are not considered to have antisense effects. At high R concentration when pA and pR strengths are low and comparable, both transcription initiation and collision rates are low, thus the truncated RNA generation rate is low, with a larger fraction located near either promoter. At moderate and high concentrations of R, *a* and *r* transcription rates are comparable. This is consistent with previously reported experimental data (Takano *et al.*, 2001). Theoretical and experimental analysis in bacteriophage 186 have shown that typically transcriptional interference in convergent promoter systems with shorter overlapping DNA is mainly due to sitting duck collisions (Callen *et al.*, 2004b; Dodd *et al.*, 2007; Sneppen *et al.*, 2005). Our simulation data also indicates that sitting duck is a major contributor to transcriptional interference in the *scbA-scbR* system.

The antisense interactions between *a* and *r* is modeled as second-order reaction (Equations 8.9-8.10). The value of rate constant k_{bar} (1×10^{-3} nM^{-1} s^{-1}) used is in the same range as reported in literature (Eguchi *et al.*, 1991; Shokeen *et al.*, 2010). The dissociation of the duplex RNA complex is assumed to follow first-order kinetics with a rate constant of $k_{\_bar}$ (Equations 8.9-8.10) (Brantl, 2007; Eguchi *et al.*, 1991; Shokeen *et al.*, 2010). The sense-antisense RNA duplex complex is assumed to be either degraded or otherwise unavailable for translation (Figure 8.1 B). The translation of ScbR and ScbA

proteins is assumed to occur only from unbound full-length transcripts, *r* and *a*, as shown by the first term in Equations 8.11 and 8.12. The binding of SCB1 to ScbR follows second-order kinetics, (Equation 8.11 and 8.13-8.14). The dissociation of the CR complex is assumed to follow first-order kinetics, denoted by $k_{\_bCR}[C_iR]$. The balance of SCB1-ScbR (C_iR) complex is given in Equations 8.14. The rate of transport of SCB1 (C_i) into or out of the cell is assumed proportional to the concentration difference of SCB1 across the cell membrane, as denoted by the $k_{se}(C_i - C_e)$ term. Degradation of C_i is considered to be a first-order process (Equation 8.13) (Mehra *et al.*, 2008).

8.3.3 Convergent transcription confers a bistable scbA/scbR genetic switch

Steady-state behavior of the *scbA-scbR* gene network was evaluated by numerically solving Equations 8.5-8.15, while keeping the extracellular SCB1 concentration (C_e) constant. A characteristic bistable response of ScbR to extracellular SCB1 is predicted, as shown in Figure 8.4. At low concentrations of SCB1 (C_e < 66 nM), the system demonstrates a single high R (OFF) state, while at high SCB1 (C_e >616 nM), the R concentration is low. Since R is a direct regulator of the downstream *cpk* gene cluster, high expression levels of R indicates an OFF state, whereas low levels of R indicates an ON state for *cpk* gene cluster. Three regions of steady states are seen. Region I (C_e <66 nM) corresponds to a stable OFF steady state with no antibiotic production. Region II corresponds to an ON state (C_e >616 nM) of antibiotic production.

The intermediate region III has two stable steady states (corresponding to either ON or OFF), in addition to one unstable (unobservable) state. Depending on the history of the system, i.e. whether the system was originally in an ON or OFF state, the trajectory of the system differs: When the system is in the initial OFF state, it continues to be in OFF state until C_e reaches 616 nM, at which point the system switches to ON (Figure 8.4). On the other hand, starting from an ON state, as C_e is decreased, the switch does not occur until a value of 66 nM is reached. Thus, in region III the system is bistable, showing alternative steady states, and therefore relatively protected against small spurious fluctuations in SCB1 concentration. Table 3 presents the range of parameters where bistability was observed.

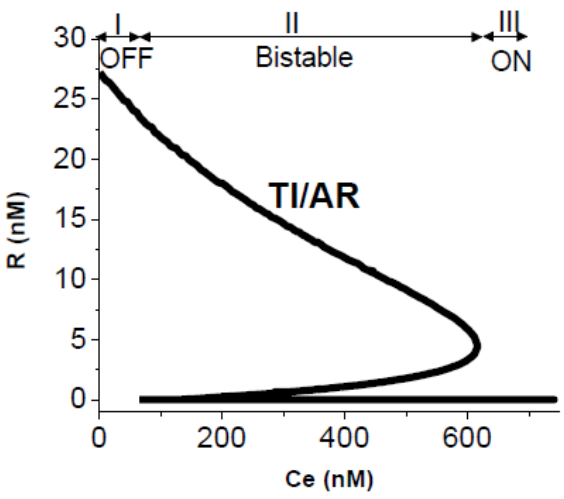


Figure 8.4: Bistable steady state response of ScbA/ScbR system to extracellular SCB1 in presence of convergent transcription. Steady state response of ScbR to constant extracellular SCB1 concentration. Parameters used for simulation are listed as nominal case in Table 3. Bistability is predicted in the case of convergent transcription. Region I and II indicate the monostable region of bistable curve corresponding to OFF and ON states respectively. Region III indicates the bistable region. The state of the system depends on its history. System originating from OFF state continues to be OFF in the bistable region, while system originating from ON state, stays ON in the bistable region. The system requires a higher concentration of SCB1 to turn ON than to turn back OFF.

8.3.4 Contribution of Transcriptional interference and antisense regulation to bistability

We evaluate the steady state behavior under the condition devoid of convergent transcription, and thus no Transcriptional interference (TI) and antisense regulation (AR) due to the removal of antisense effects between full-length *a* and *r* RNA (Figure 8.5 A, TI^-/AR^- case). TI effect is eliminated from the model by assigning rates of production of full-length transcripts k_a and k_r equal to the RNAP firing rate from pA and pR, i.e. $k_a = k_{pA}$ and $k_r = k_{pR}$. The AR effect is removed by setting the parameters $k_{bar} = k_{\_bar} = k_{dar} = 0$. Without AR and TI, the system loses bistability (Figure 8.5 A, TI^-/AR^- case). A broad scan of the parameter space did not yield a bistable response for any selection of parameters (Table 8.3).

We next evaluated the relative contribution of TI and antisense regulation to the bistable response. Removing only antisense regulation but not TI, by setting the parameters $k_{bar} = k_{\_bar} = k_{dar} = 0$ eliminated the bistable behavior (Figure 8.5 A, TI/AR^- case). With antisense regulation alone without TI effect ($k_a = k_{pA}$ and $k_r = k_{pR}$) the system could still demonstrate bistability (Figure 8.5 A, TI^-/AR case). Antisense regulation is thus critical for the bistable response in this system. Interestingly, when both TI and AR effects are considered, the system requires a lower value of R to transition into the ON state compared to the case of AR alone, indicating that the combined effect imposes a slightly higher suppression of R.

To assess the contribution of TI to the bistability, we simulated conditions with varied TI effects by ranging the RNAP initiation time τ (i.e. the time taken by a closed promoter complex to transition into an elongation complex) between 0s (very fast) to 10 s (very slow). Under very low concentrations of R (such as [R] tending to 0 nM), for τ=5s and 10s, the time taken for RNAP to initiate at the promoters is longer than corresponding RNAP binding time intervals (2.2 s and 4.8 s at promoters pA and pR respectively for [R] =0 nM). In such a situation, the model considers RNAP does not bind at a promoter till the promoter is cleared. This is achieved in the simulations by incrementing the n^{th} round of RNAP binding at a promoter to n+1 in case promoter is not cleared. As expected, increasing τ, results in higher frequency of sitting duck collisions and consequently decrease in the transcription rate of *a*, *r* (Figure 8.6 A-B).

We noticed that steady state response of R to C_e in presence of TI effect only showed a ramping-down switch response only for high values of τ (Figure 8.5 B). For low values of τ, the system continued to stay at an OFF state even at high levels of SCB1 induction. However, bistability is never achieved in presence of TI alone (Figure 8.5 B). When both TI and antisense effects (TI/AR) are considered, as τ increases, the bistable region decreases and at high values of τ bistability is lost. At high τ, the transcription rates of full-length *a* and *r* RNA decrease, resulting in weakening of antisense effects. Thus, decreasing the relative contribution of antisense effect to TI effect causes loss of bistability, indicating that the non-linear effect offered by antisense interactions between the full-length RNA is essential for bistability (Figure 8.5 C).

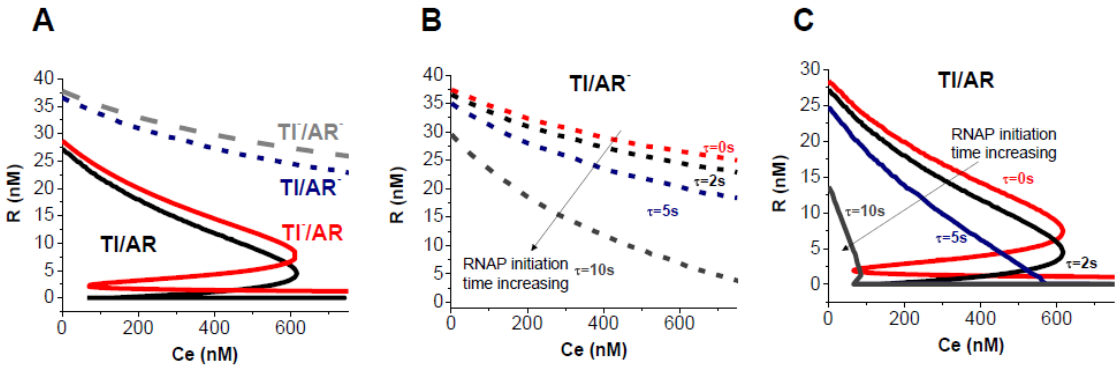


Figure 8.5: Relative contribution of RNAP collision and antisense regulation effects to bistable switch response. (A) Bistability is lost in the absence of TI and antisense RNA interaction effects (TI⁻/AR⁻). In presence of only TI effects (TI/AR⁻), the bistable behavior is lost, system ceases to behave as a switch. Bistability is restored when only antisense RNA interaction between full-length RNA is considered in absence of RNAP collision (TI⁻/AR). The system exerts tighter regulation on ScbR expression in presence of convergent transcription (TI/AR). **(B)** Steady state expression level of repressor R to extracellular SCB1 concentration in presence of TI effect only for different RNAP initiation time (τ) at promoter pA and pR. **(C)** Steady state expression level of repressor R to extracellular SCB1 concentration in presence of both TI and AR effects for different RNAP initiation time (τ) at promoter pA and pR.

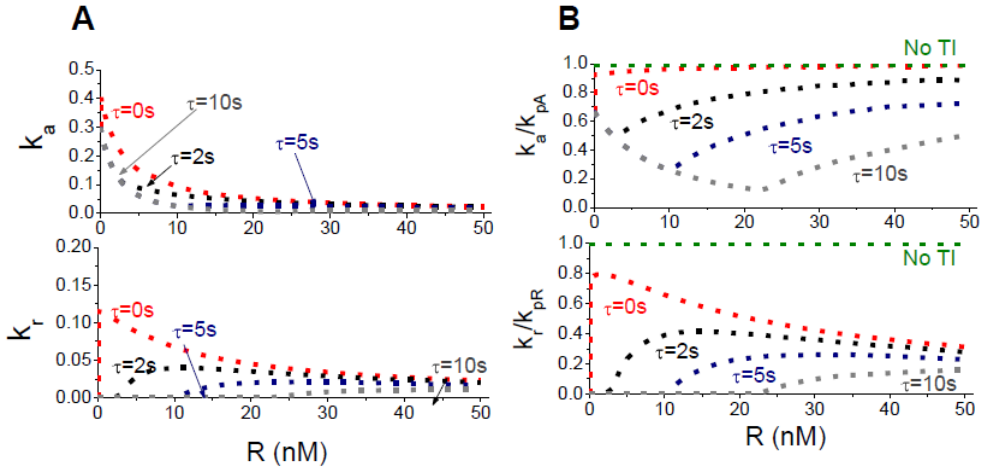


Figure 8.6: Effect of different RNAP initiation time (τ) on the resultant transcriptional interference. (A) Rate of transcription of full-length RNA (k_r and k_a) and (B) rate of transcription of full-length RNA normalized to RNAP binding rates (k_r/k_{pR} and k_a/k_{pA}), for different concentrations of repressor ScbR, for different RNAP initiation time (τ) at promoters pR and pA.

Table 8.3 Parameter values and their range for which bistability is observed

Parameter	Description	Estimated value for bistability	Range of bistability		Remarks/ Reference	Units
			Min. evaluated	Max. evaluated		
K_{OR}	Equilibrium binding constant of ScbR to O_R	8.82	0.44	37.92	(Santillan and Mackey, 2001; Wong *et al.*, 1997)	nM
K_{OA}	Equilibrium Binding constant of ScbR to O_A	2.68	0.12	4.92	(Santillan and Mackey, 2001; Wong *et al.*, 1997)	nM
$k_{pR-\max}$	Transcription from P_R in de-repressed state	1.5×10^{-1}	7.5×10^{-3}	3.0×10^{-1}	(Vo *et al.*, 2003)	s^{-1}
$k_{pA-\max}$	Transcription from P_A in de-repressed state	4.5×10^{-1}	2.25×10^{-2}	9.0×10^{-1}	(Vo *et al.*, 2003)	s^{-1}
$k_{pR-\min}$	Transcription from P_R in repressed state	1.0×10^{-3}	5.0×10^{-5}	3.0×10^{-3}	(Vo *et al.*, 2003)	s^{-1}
$k_{pA-\min}$	Transcription from P_A in repressed state	8.0×10^{-4}	4.0×10^{-5}	2.8×10^{-3}	(Vo *et al.*, 2003)	s^{-1}
k_{dr}	Degradation of full-length *r* RNA	7.0×10^{-3}	3.5×10^{-4}	1.4×10^{-2}	(Bernstein *et al.*, 2002)	s^{-1}
k_{da}	Degradation of full-length *a* RNA	8.1×10^{-4}	4.1×10^{-4}	1.6×10^{-2}	(Bernstein *et al.*, 2002)	s^{-1}
k_R	ScbR protein translation	3.6×10^{-1}	9.0×10^{-2}	3.8×10^{-1}	(Voigt *et al.*, 2005; Wong *et al.*, 1997)	s^{-1}
k_A	ScbA protein translation	6.6×10^{-2}	1.0×10^{-2}	1.2×10^{-1}	(Voigt *et al.*, 2005; Wong *et al.*, 1997)	s^{-1}
k_{dR}	ScbR protein degradation	4.0×10^{-3}	1.0×10^{-3}	8.0×10^{-3}	(Wong *et al.*, 1997)	s^{-1}
k_{dA}	ScbA protein degradation	1.8×10^{-3}	8.0×10^{-4}	2.0×10^{-3}	(Wong *et al.*, 1997)	s^{-1}
μ	growth rate	6.0×10^{-5}	6.0×10^{-5}	8.0×10^{-5}	(Cox, 2004)	s^{-1}
k_C	SCB1 synthesis	7.4×10^{-1}	7.4×10^{-2}	37	(Parsek *et al.*, 1999)	s^{-1}
k_{dC}	SCB1 degradation	6.7×10^{-5}	6.7×10^{-6}	6.7×10^{-3}	Max. half life 1 hr.	s^{-1}
k_{se}	SCB1 secretion	8.3×10^{-2}	8.3×10^{-2}	4.2	(Goryachev *et al.*, 2005)	s^{-1}
k_{bCR}	Binding of ScbR and SCB1 to form SCB1-ScbR complex	8.3×10^{-2}	4.2×10^{-3}	2.53×10^{-1}	(Santillan and Mackey, 2001)	nM^{-1} s^{-1}
$k_{\_bCR}$	Unbinding of SCB1-ScbR complex	1.7×10^{2}	8.5	1.95×10^{2}	(Santillan and Mackey, 2001)	s^{-1}
k_{dCR}	SCB1-ScbR degradation	3.1×10^{-3}	3.1×10^{-3}	6.8×10^{-2}	(Santillan and Mackey, 2001)	s^{-1}
k_{bar}	Binding rate constant of full-length (*a, r*) to form hybrid RNA complexes *a:r*	1.0×10^{-3}	6.5×10^{-4}	1.6×10^{-1}	(Eguchi *et al.*, 1991)	$nM^{-1}s^{-1}$
$k_{\_bar}$	Unbinding of *ar* complex	1.0×10^{-2}	0	2.0×10^{-1}	(Eguchi *et al.*, 1991)	s^{-1}
k_{dar}	RNA hybrid complex *ar* degradation	1×10^{-2}	0	1.24	(Eguchi *et al.*, 1991)	s^{-1}

8.3.5 Sensitivity of the bistable response

We evaluated how the steady state behavior of ScbA/R system is affected by a number of parameters critical to the observed bistability. The key outcome of antisense regulation is the decrease in *a* and *r*. We evaluated the effect of antisense interaction by varying the value of k_{bar}, while keeping the values of the remaining parameters constant. As k_{bar} is decreased, the bistable region diminishes (Figure 8.7 A). Eventually, the bistable response becomes a ramping response. Increasing k_{bar} caused the system to shift from reversible bistable behavior to one that is irreversible with the bistable region extending to the zero concentration of C_e. A system that originates from an ON state (low ScbR) will remain ON with decreasing C_e. This is explained by the fact that high rates of binding of sense and antisense RNA (high k_{bar}) will sequester free *r* RNA. As a result the system will not be able to produce enough repressor R to allow the system to return back to the OFF condition.

In *S. coelicolor* cultures, SCB1 and antibiotic production commence during the transition phase from a rapid growth to a slower growth. We evaluated the effect of growth rate on bistability. Increasing the growth rate causes the bistable curve of the response of R to SCB1 to shift to the right; however, the bistable region shrinks and ultimately vanishes at very high values of the specific growth rate (Figure 8.7 B). At 10 fold decreased growth rate than the nominal values used in Table 8.3, the system demonstrates irreversible bistability. This implies, that at slower growth rates such as in stationary phase, once the switch is ON, the system continues to stay ON, the cells will

continue the production of antibiotics and/or secondary metabolites, until the growth rate increases again.

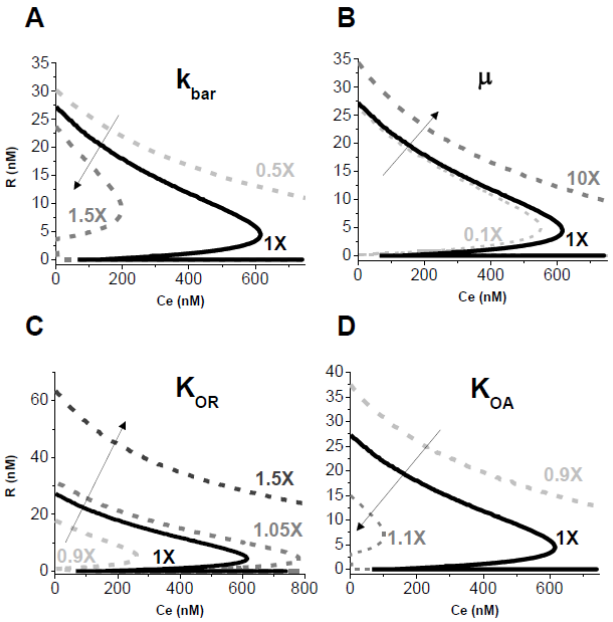


Figure 8.7: Effect of single-parameter perturbation on steady state response of ScbA-ScbR system to constant extracellular SCB1. Results show the effect of varying one (indicated) parameter while keeping the rest constant at the nominal values listed in Table 3. The solid black line (1X) in each plot corresponds to the nominal parameter values. Effect of changing parameters: **(A)** RNA: RNA interaction rate constant, **(B)** growth rate and **(C-D)** equilibrium rate constant corresponding to binding of repressor R to operator sites O_R (C) and O_A (D) on the steady state response of ScbA-ScbR gene network.

8.3.6 Effect of ScbR repression on promoters pA and pR

The strength of promoters pA and pR is critical to the bistability. Since the repressor R is a direct effector of the promoter strength, we next evaluated the effect of ScbR repression on promoters pR and pA. The repressive effect of R on promoters pA and pR is characterized by the equilibrium binding rate constant of R at the operator sites O_R and O_A, K_{OR} and K_{OA}, respectively. Higher K_{OR} and K_{OA} implies lower binding affinity of R towards operators O_R and O_A, respectively, hence correspondingly a lower repressive effect on promoters pR and pA. The effect of changing K_{OR} on the response of

R to SCB1 is shown in Figure 8.7 C. In this case, decreasing K_{OR} results in an irreversible bistable switch, implying tighter regulation of pR by R actually prevents the system from turning OFF. Also, the steady state level of R in the OFF state is lower than the nominal case. The tighter repression of R causes a lower expression level of R: once the system is ON, the system is never able to switch OFF, as the excess SCB1 sequesters the free R. Increasing K_{OR}, i.e. decreased repression, causes the bistable curve to shift to the right along with widening of the bistable region, implying that a higher SCB1 concentration is required to induce the system. Changing the rate constant K_{OA} shows an effect opposite to that of K_{OR} (Figure 8.7 D). When increasing K_{OA}, thus decreasing binding affinity of R to O_A, the system begins to demonstrate irreversible bistability. On the other hand, with decreasing K_{OA}, the bistable region expands considerably, such that the system turns ON only at much higher SCB1 concentration.

8.4 Discussion

Cellular decisions are mediated through genetic switches which arise from interactions between simple biological molecules. Robust genetic switches often demonstrate bistability, which implies that the system exists only in two discrete states (Chatterjee *et al.*, 2008), i.e. cells either exist in an ON state or an OFF state. Several key physiological decisions such as the transition between lysogeny and lytic state in bacteriophage λ (Tian and Burrage, 2004), determining competence in *Bacillus subtilis* (Maamar and Dubnau, 2005), or between differentiation and self-renewal in stem-cells

(Chickarmane *et al.*, 2006a; Wang *et al.*, 2009), have been shown to be characterized by bistable behavior. Bistable systems usually demonstrate hysteresis, making them less susceptible to fluctuating noise around the decision point, as the threshold required for the system to switch from OFF to ON is different from that going back from ON to OFF.

S. coelicolor switches from a vegetative growth state (antibiotic production OFF) to a stationary state (antibiotic production ON) (Mehra *et al.*, 2008c; Takano *et al.*, 2001). The decision to switch from OFF to ON is a critical one, as secondary metabolites such as antibiotics can be toxic to the producers themselves. The switch from the vegetative to the antibiotic-producing state is triggered when the concentration of extracellular SCB1 reaches a critical threshold. Previously, we have shown that formation of a hypothetical ScbA-ScbR protein complex acting as a positive regulator of transcription from pA (Mehra *et al.*, 2008), would be required for bistable switching behavior. Removal of the ScbA-ScbR complex from the model resulted in loss of the switch response as demonstrated by the TI^-/AR^- case in the current analysis (Figure 8.5 A).

In this work we have shown that convergent transcription in the *scbA-scbR* locus can restore the bistable switch behavior in the absence of ScbA-ScbR protein complex. Interestingly, we found that convergent transcription gave rise to bistable behavior over a wider parameter range compared to the earlier model that depended on an ScbA-ScbR protein complex (Mehra *et al.*, 2008). This robustness is an implicit argument in favor of the convergent transcription mechanism (Morohashi *et al.*, 2002; Stelling *et al.*, 2004). The bistable behavior in the ScbA/ScbR system appears to be robust as it is relatively

insensitive to the value of the parameters used. Varying the value of the parameter both individually (Figure 8.6, Figure 8.8 and Figure 8.9) and in combination with other parameters we examined the maximum range of system bistability for each parameter (Table 8.3). When parameters were varied in combination with other parameters, the system was found to demonstrate bistability in a range of parameter values by at least tenfold.

Convergent transcription gives rise to two mechanisms of gene regulation: transcriptional interference and antisense interaction. Transcriptional interference has been shown to play a regulatory role in the expression from pR-pL promoter pair in bacteriophage 186 (Callen *et al.*, 2004), P_Q-P_X promoter pair controlling conjugation of pCF10 plasmid in *Enterococcus faecalis* (Chatterjee *et al.*, 2011), S-box antisense RNA repression of *ubi-mccBA* mRNA of *C. acetobutylicum* (Andre *et al.*, 2008) and in cell fate control between diploid and haploid states in *IME4* locus of *Saccharomyces cerevisiae* (Hongay *et al.*, 2006). In ScbA-ScbR system, a single repressor ScbR regulates expression from both promoters. At high levels of repressor ScbR, expression from promoter pR and pA are comparable (Figure 8. 3A-C). At low free R conditions, transcription rate from the pA promoter is higher than from the pR promoter. pA thus functions as the dominant promoter and its transcription suppresses transcription from pR mainly via sitting duck collisions and occlusion of pR promoter. RNAP collision has two effects on transcription: first, decreasing the expression of full-length transcripts from both promoters; second, the effect of TI is more severe for transcription from the weaker

promoter, thus resulting in the amplification of the difference of the levels of the transcripts from the two promoters.

Transcriptional interference may have a more pronounced effect in systems with a longer overlapping region and in the presence of pause sites within the overlapping region (Callen *et al.,* 2004b; Chatterjee *et al.,* 2011; Palmer *et al.,* 2009). In the case of PR-PRE promoter pair in bacteriophage λ, the presence of pause sites in the 320 bp of overlapping DNA enhanced the effect of transcriptional interference (Callen *et al.,* 2004; Palmer *et al.,* 2009). In our simulations, increasing the time required for transcriptional initiation after RNAP binds at promoter pA and pR from 2s to 10s, increased the rate of sitting duck collisions resulting in further decrease in the production of full-length RNA (Figure 8. 6 A-B).

Antisense interaction is the second layer of regulation offered by convergent transcription. Full-length transcripts have complementary counterparts from the opposing promoters which may elicit antisense interactions. The resulting hybrid RNA complex are subjected to degradation or rendered untranslatable (Brantl, 2002; Brantl, 2007). In this system we assume RNA interaction tends to sequester full-length transcripts, which are capable of being translated. The effect is again more pronounced for the less abundant RNA, as the depletion effect is more pronounced.

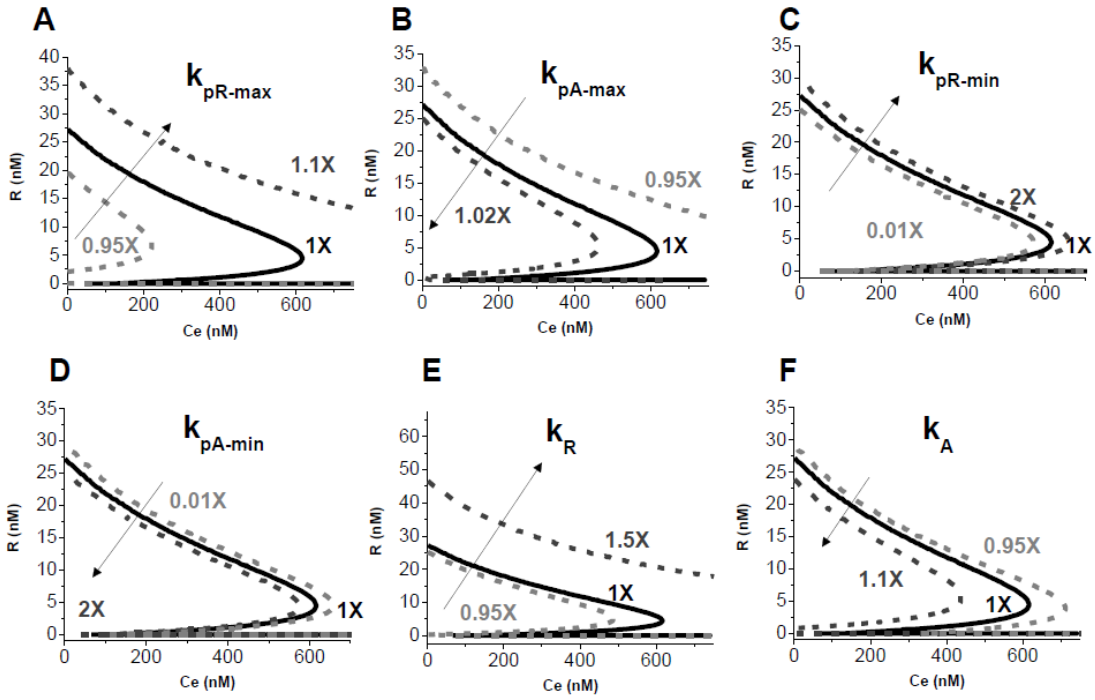


Figure 8.8: Effect of single-parameter perturbation for parameters describing transcription and translation on the steady state response of ScbA-ScbR system to constant extracellular SCB1. Results show the effect of varying one (indicated) parameter while keeping the rest constant at the nominal values listed in Table 3. The solid black line (1X) in each plot corresponds to the nominal parameter values. The parameter being varied include, transcription rate constants: (A) $k_{pR\text{-}max}$ (B) $k_{pA\text{-}max}$ (C) $k_{pR\text{-}min}$ (D) $k_{pA\text{-}min}$, translation rate constants: (E) k_R (F) k_A.

Similar role of antisense RNA in down regulating synthesis of proteins from *cis*-encoded genes has been reported in other systems, including *Sok* mRNA of plasmid R1 in *E.coli* (Thisted and Gerdes, 1992), antisense RNA *alr1690-furA* regulating expression of transcriptional repressor FurA in cyanobacterium *Anabena* sp. PCC 7120 (Hernandez *et al.*, 2006) and other systems reviewed in (Thomason and Storz, 2010). The repressive effect of antisense transcripts has also been experimentally shown in the convergent

promoter system of *prgQ-prgX* gene pair present on plasmid pCF10 in *Enterococcus faecalis* (Chatterjee *et al.*, 2011). Antisense-RNAs have been reported to play a regulatory role in other bacterial systems, including long antisense RNAs such as 2200 nt *alr1690-α-furA* antisense RNA regulating *furA* transcript in the cyanobacterium *Anabaena* sp. strain PCC 7120 (Hernandez *et al.*, 2006), 1200 nt *AmgR* RNA encoded convergent to *mgtCBR* operon in *S. enterica* (Lee and Groisman, 2010), and relatively shorter antisense RNAs such as 108 nt RNAI RNA controlling copy number of plasmid ColEI (Tomizawa and Itoh, 1981), 69 nt Sar RNA of bacteriophage 22 repressing Ant protein (Liao *et al.*, 1987), 77 nt OOP RNA of bacteriophage λ repressing CII protein (Krinke and Wulff, 1987), 90 nucleotide CopA RNA controlling copy number of plasmid R1 (Stougaard *et al.*, 1981) and 70 nt Sok RNA repressing post-segregation killing in plasmid R1 (Thisted and Gerdes, 1992).

Expression of sense and antisense full-length *a* and *r* RNA has been reported previously (Mehra *et al.*, 2008c; Takano *et al.*, 2001). Under repressed conditions, *scbR* and *scbA* transcript levels have been shown to be similar, whereas, under induced conditions *scbA* transcript levels exceed *scbR* (Takano *et al.*, 2001). These are consistent with the results of the mathematical model (Figure 8.3 A-C). Moreover, recent work indicates presence of large number of *cis* antisense RNA in the model streptomycete *Streptomyces coelicolor* (D'Alia *et al.*; Panek *et al.*, 2008; Swiercz *et al.*, 2008).

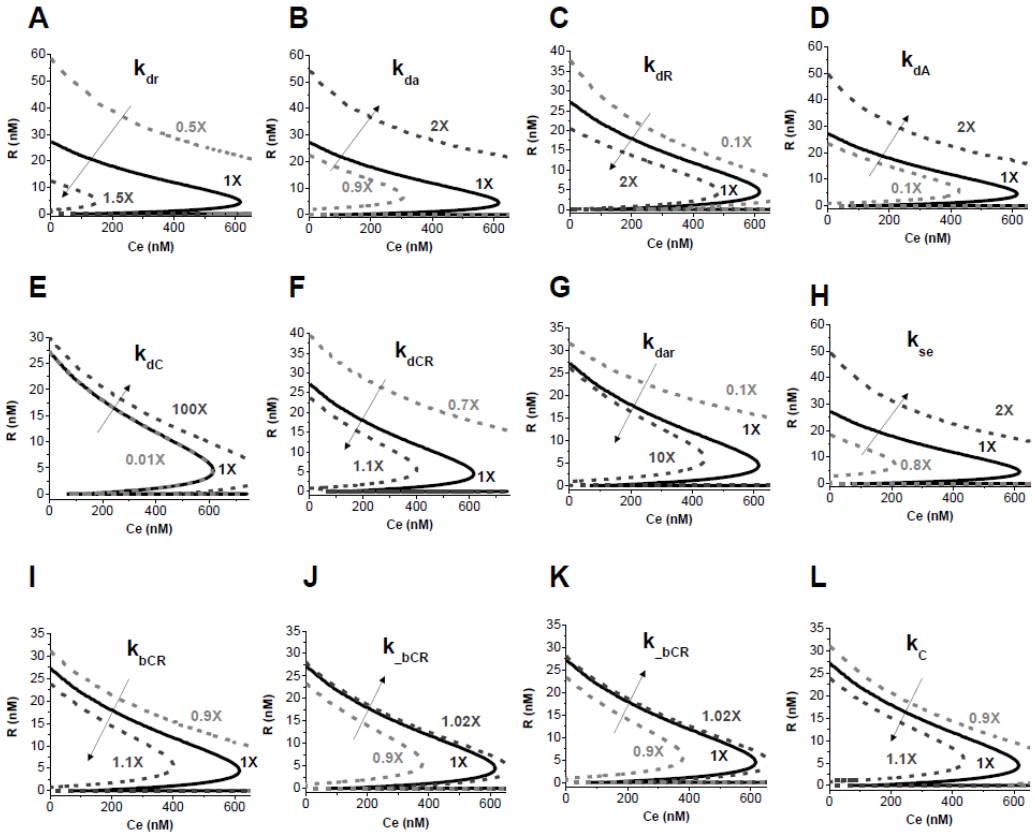


Figure 8.9: Effect of single-parameter perturbation on steady state response of ScbA-ScbR system to constant extracellular SCB1. Results show the effect of varying one (indicated) parameter while keeping the rest constant at the nominal values listed in Table 3. The solid black line (1X) in each plot corresponds to the nominal parameter values. The parameter being varied include degradation rates: (A) k_{dr} (B) k_{da} (C) k_{dR} (D) k_{dA} (E) k_{dC} (F) k_{dCR} (G) k_{dar}, (H) SCB1 secretion rate: k_{se}, (I) SCB1-ScbR binding rate constant, k_{bCR} , Unbinding rate constants: (J) $k_{\_bCR}$ (K) $k_{\_bar}$ and (L) SCB1 production rate, k_C.

In the *scbA-scbR* system, due to the short overlapping distance of 53 bp, RNAP collision would generate very short sizes of truncated RNA, which would lack the structural features required for exerting antisense effects. In systems which have a longer length of overlapping DNA, TI could potentially give rise to two-fold regulation, (i) reduction of full-length transcript levels, (ii) generation of truncated RNA with reasonable structural features to exert antisense interaction on counter transcripts. Presence of truncated RNA during convergent transcription have been reported in a few bacterial systems with longer overlapping DNA, such as in the 1.2 kb overlap of *mgtCBR-AmgR* locus in *S. Enterica* under conditions of low Mg^{2+} ions when both convergent promoters were induced by transcriptional activator PhoP (Lee and Groisman, 2010). Convergent transcription within the 223 bp overlapping DNA of the *prgQ-prgX* gene-pair controlling conjugation of plasmid pCF10 in *Enterococcus faecalis*, driven by convergent promoters P_Q-P_X respectively, has been shown to produce increased-levels of truncated RNA within the overlapping locus when promoters are of similar strengths (Chatterjee *et al.*, 2011). Interestingly, in this system truncated RNA in both sense and antisense orientation were shown to participate in antisense interactions with counter transcripts.

In a convergent promoter pair the effect of RNAP collision on successful transcription is asymmetrical to the two promoters even though the frequency of collision is the same for both. When the strength of the two promoters are substantially different, the collision loci will shift toward the weaker promoter, conversely generating long

truncated transcripts of the stronger promoter which can exert further antisense effect on the weak transcript, enhancing contrast between the effect of strong and weak promoters. The consequence of convergent transcription is thus highly non-linear in terms of the loci of RNAP collision and consequently the antisense effects, and the response to induction or repression. However, antisense interaction offers a higher degree of non-linearity (second order reaction) compared to TI and thus is essential for the predicted bistability in the *scbA-scbR* system (Figure 8.5 A).

Convergent transcription in the *prgQ-prgX* gene-pair in *E. faecalis* has been shown to attribute a bistable genetic switch controlling transfer of plasmid pCF10 via conjugation (Chatterjee *et al.*, 2011). Similarly, convergent transcription in the *furA-alr1690 mRNA* locus allows for regulation of FurA, a prokaryotic transcriptional factor that is involved in regulation of Ferric uptake during environmental stress response (Hernandez *et al.*, 2006). In the *mgtCBR-AmgR* locus, the expression of antisense *AmgR* RNA was shown to diminish expression of MgtC and MgtB proteins, while the inactivation of *AmgR* promoter resulted in increased virulence in mice (Lee and Groisman, 2010).These examples clearly highlight regulatory potential of convergent transcription and its role in key biological decision making.

Recent work has led to the discovery of *cis* aRNA in *Streptomyces coelicolor* (D'Alia *et al.*; Panek *et al.*, 2008; Swiercz *et al.*, 2008). Nearly 3600 *cis* non-coding RNA have been predicted, some of these that were experimentally validated show differential expression under certain growth conditions (D'Alia *et al.*, 2009). For example, over

expression of chromosomal *cis* non-coding RNA *cnc2198.1* found antisense to glutamine synthetase I has been shown to result in decrease in protein expression, growth and production of antibiotics in *S.coelicolor* (D'Alia *et al.*, 2009). Bioinformatic approaches have also been used to predict novel *cis*-asRNA's (Panek *et al.*, 2008; Swiercz *et al.*, 2008), some of which have been experimentally validated and shown to be regulated developmentally (Swiercz *et al.*, 2008).

Given the widespread presence of convergent transcription in both prokaryotic and eukaryotic organisms, it is highly plausible that the arrangement of gene pairs in such organization confers some biological regulatory function. It has been reported that 1947 such convergent promoter pairs are present in the mouse genome and transcriptome analysis provides evidence that a significant fraction of these have asymmetrical transcriptional regulation (Kiyosawa *et al.*, 2005). The *Streptomyces coelicolor* genome consists of 1429 pairs of divergently transcribed genes, however, transcript start sites have been determined for only a couple of cases (http://streptomyces.org.uk/). Based on known open reading frames, at least 80 gene pairs are arranged in convergent orientation, although the actual extent of transcript overlap is not known. The asymmetry in transcription rates change upon induction (de-repression), analogous to the ScbA-ScbR system discussed in this work, could operate in many of the convergent promoter pairs present in *S. coelicolor*. It is possible that mechanisms of transcriptional interference and antisense regulation operate in these convergent promoter systems and play regulatory

roles in gene expression. Despite its structural simplicity, convergent gene-pairs may harbor some regulatory complexity yet to be fully investigated and exploited.

8.5 Materials and Methods

8.5.1 Steady state and dynamic analysis of mathematical model

Numerical solutions to the ordinary differential equations were solved using the stiff differential equations solvers ode23s in Matlab®. The steady states for the equations were computed in Mathematica. The stability of solutions obtained was characterized by Eigenvalues of the Jacobian. The complete set of kinetic parameters involved in the above model is listed in Table 3. The range of values for each parameter in Table 8.3 was obtained from the literature (Bernstein *et al.*, 2002; Cox, 2004; Goryachev *et al.*, 2005; Parsek *et al.*, 1999; Santillan and Mackey, 2001; Vo *et al.*, 2003; Voigt *et al.*, 2005; Wong *et al.*, 1997). The parameter range was explored to determine the capability of such a system to show the desired system dynamics. *In silico* structures of RNA were generated using Sfold software (Ding and Lawrence, 2003).

CHAPTER 9

SUMMARY AND CONCLUDING REMARKS

Increased prevalence of antibiotic resistance within bacteria, which has been accelerated due to resistance transfer via mechanism of conjugation, is a growing challenge for future therapy of bacterial infections (Tendolkar *et al.*, 2003). The problem is circumvented by the fact that prolonged exposure to antibiotics increases the risk of antibiotic resistance. As antibiotic resistance becomes common, a greater need for alternative therapies is needed. A key to solving the antibiotic resistance problem is to understand the mechanism of resistance transfer and target those processes which control the transfer to develop new therapy.

In this thesis, we investigated pheromone-induced transfer of antibiotic resistance plasmid pCF10 in *Enterococcus faecalis* via conjugation between donor cells, which contains tetracycline resistance conferring plasmid pCF10, and recipient cells, which lack this plasmid. We studied the gene regulatory components of the genetic switch controlling the onset of conjugation using both mathematical and experimental approaches to (i) identify the key regulatory elements and (ii) perturb genetic switch controlling conjugation as a strategy to counter drug resistance transfer.

A set of fundamental aspects pertaining to regulatory mechanisms during convergent transcription have been elucidated in Chapter 4-5 and 8 (Chatterjee *et al.*,

2011b; Chatterjee *et al.*, 2011). A key topic of the current research included studying the biological function of convergent (transcription with focus on the regulatory role of RNA polymerase traffic (Transcriptional Interference) and antisense RNA, especially its role in controlling transfer of plasmid pCF10 in bacterium *Enterococcus faecalis*, a nosocomial pathogen. Deterministic mathematical models based on a reduced genetic-network based on pCF10 system were developed (Chapter 5). Mathematical model guided experimental design led to discovery of a novel mechanism in the control of transfer of this plasmid between donor and recipient cells (Chapter 4). We identified reciprocal regulation of sense and antisense RNA produced as a result of convergent transcription in this system. Using both experimentation and rigorous mathematical modeling, we found that antisense transcription confers the pCF10 system a robust bistable-switch like behavior in controlling antibiotic resistance transfer (Chatterjee *et al.*, 2008; Chatterjee *et al.*, 2011).

This work has two implications. Firstly, given that convergent gene pairs are wide spread in both eukaryotic (thousands) and prokaryotic genomes (hundreds) and are speculated to be involved in gene regulation, the results from our work have implications for other biological systems that show similar gene arrangement such as those involved in HIV latency and expression of cancer related genes. Secondly, this work hints towards potential future therapy that would prevent drug resistance transfer. For example, a molecule that could potentially target this genetic switch and maintain it at an 'OFF' state, in conjunction with antibiotics, would be toxic to bacteria.

The effect of pCF10 plasmid copy number on efficiency of antibiotic resistance transfer in *E. faecalis* was studied both from a mathematical and experimental perspective (Chapter 6) (Cook *et al.*, 2011). Model prediction showed that by varying pCF10 copy number, donor cells can adjust their response to induction, such that cells with higher pCF10 copy number require a high level of induction to turn ON conjugation. Experimental data fully complemented model prediction, with cells growing under natural biofilm condition demonstrating an average high copy of pCF10 plasmid, as well as the fact that OFF donor population possessed on an average higher plasmid copy number. Though it is counter-intuitive that higher copies of plasmid result in lower degree of plasmid transfer, a higher threshold for induction allows the donors cells to ensure that the inducer signal is true and that there is a neighboring recipient cell. Such a behavior is specific to pCF10 and is attributed to the negative regulatory components present on pCF10 plasmid. The fact that by tweaking the number of copies of pCF10 plasmid, donor cells can control resistance transfer and introduce population heterogeneity, is an important consideration towards development of future drug therapies to treat bacteria in their natural environment.

A natural follow-up of the study on bistable *prgQ-prgX* gene-switch (Chapter 4-5), led to questions as to how donor cells control turning-off of conjugation (Chapter 7). Additionally, discovery of higher- pCF10 copy number in biofilm donor cells pointed towards that fact that donor cells prefer to stay in “OFF” state to avoid false signal (Chapter 6). In Chapter 7, we described a novel role for iCF10, the inhibitor signaling

peptide for pCF10 based system (Buttaro *et al.*, 2000), as a quorum-sensing signal with the aim of turning-OFF conjugation at a population-wide level. Unlike conventional quorum sensing systems, where the signaling molecule turns-ON gene expression at a population scale (Waters and Bassler, 2005), the pCF10 system is unique in the sense that the quorum sensing signal coordinates turning-OFF of conjugation related gene expression among the donor population. Both mathematical modeling and experimental analysis of dynamic response of pCF10 to induction indicate that this negative feed-back loop causes a brief "pulse-like" expression of the entire operon (Chatterjee *et al*, to be submitted). Such an interplay of positive and negative feedback loops allows donor cells to quickly transition between ON and OFF states, which may be critical both for the transfer of plasmid and survival of the donor cell.

Besides *E. faecalis*, using mathematical modeling the regulatory role of convergent transcription in controlling genetic switch for onset of antibiotic production in bacterium *Streptomyces coelicolor* was also investigated (Chapter 8) (Chatterjee *et al.*, 2011). Given that *S. coelicolor* has a large number of convergent promoters (approximately 1947) and that a large fraction of these have not yet been characterized, the mechanisms of transcriptional interference and antisense RNA interaction described in this thesis may be operative in these multiple loci on the *S. coelicolor* chromosome.

Future work would include developing understanding the turning-OFF mechanism of conjugation. The response of donor cells to different concentrations of iCF10 and cCF10 would be evaluated. To evaluate this, the ΔEep mutant would be used

to estimate the response of Q_L RNA to various combinations of iCF10 and cCF10. In our study to date Q_L has been considered as the state variable because it is indicative of the synthesis of the conjugation machinery. Promoter P_Q has been thought to lead to a very long transcript of approximately 20 kb in length, of which Q_L is a processed product. In addition to promoter P_Q other transcription initiation or termination sites may exist along the coding region of the long transcript but have not been uncovered. To investigate the expression of downstream genes on pCF10, RNA-seq and quantitative shotgun proteomics would be performed to examine the dynamics of conjugation machinery.

RNA from induced cells at various time points (0, 30, 60 and 120 minutes) post-induction will be used for RNA-seq with 5' end triphosphate tagging and without to identify transcription initiation sites and Tobacco acid Pyrophosphatase to convert 5' tri-phosphate to mono-phosphate. A sharp drop in sequencing depth at specific sites would be indicative of post-transcriptional processing, while additional sequences with 5' end phosphate labeling would indicate additional transcription initiation. The data will also provide the kinetics of turning ON and OFF of all genes downstream that may unveil hidden regulation of conjugation not known previously. The dynamics of transcript expression would be compared with the iTRAQ proteomics data discussed in Chapter 7. So far, we performed proteomics runs under induced condition and quantified about half of the downstream proteins present on pCF10. More fractions from cation exchange column will be used for second dimension of LC; the theoretical peptide spectra from coding sequences will be used for targeted proteomics to identify more proteins.

Comparison of RNA-Seq data and proteomics data would provide insight into regulation post-induction and hint towards mechanisms involved in turning-OFF of conjugation.

The study of antisense RNAs is one of the most exciting areas of regulatory RNA research and has tremendous potential as a candidate for future drug therapy (Thomason and Storz, 2010). Future applications include discovery of novel antisense RNAs and synthetic gene-networks for engineered-antisense RNA regulation. With the recent advances in transcriptional analysis, especially with the advent of RNA-Sequencing technology, hundreds of antisense RNAs are being reported, highlighting the tremendous scope of research in this area. Some questions that need to be asked include, are convergent gene pair's part of regulatory circuit and demonstrate switch behavior via mechanisms similar to those reported in this thesis. We are currently performing RNA-Seq on RNA samples obtained from *S. coelicolor* and *E. faecalis* grown under various growth conditions to ascertain the extent of convergent transcription in these two very different microorganisms.

The presence of multiple layers of regulation that control conjugative transfer of plasmid pCF10 in *E. faecalis*, make pCF10 system a very intriguing system of study. This also reveals multiple targets for potential drug therapy. The work in this thesis directly impacts the current problem of antibiotic resistance in bacteria, especially because *E. faecalis* is currently one of the three most frequent causes of hospital-acquired infections and constitutes a major problem for the current health care system (Tendolkar *et al.*, 2003). In this work, the combined modeling and experimental approaches provide

a powerful model of drug resistance transfer in *E. faecalis.* The information generated will allow us to identify potential targets for perturbing the genetic network controlling drug-resistance transfer in *E. faecalis*. The ultimate goal of the future research would be to identify and experimentally test such targets that can lead to cessation of antibiotic resistance transfer in *E. faecalis* and thus serve as a model for other drug resistance transferring bacterial systems.

CHAPTER 10

BIBLIOGRAPHY

(2000) World Health Organization report.

Aakra, A., Vebo, H., Snipen, L., Hirt, H., Aastveit, A., Kapur, V., Dunny, G., Murray, B., and Nes, I.F. (2005) Transcriptional response of Enterococcus faecalis V583 to erythromycin. *Antimicrobial Agents and Chemotherapy* **49**: 2246-2259.

Ajo-Franklin, C.M., Drubin, D.A., Eskin, J.A., Gee, E.P.S., Landgraf, D., Phillips, I., and Silver, P.A. (2007) Rational design of memory in eukaryotic cells. *Genes Dev.* **21**: 2271-2276.

An, F.Y., and Clewell, D.B. (2002) Identification of the cAD1 sex pheromone precursor in Enterococcus faecalis. *Journal of Bacteriology* **184**: 1880-1887.

Andre, G., Even, S., Putzer, H., Burguiere, P., Croux, C., Danchin, A., Martin-Verstraete, I., and Soutourina, O. (2008) S-box and T-box riboswitches and antisense RNA control a sulfur metabolic operon of Clostridium acetobutylicum. *Nucleic Acids Research* **36**: 5955-5969.

Antiporta, M.H., and Dunny, G.M. (2002) ccfA, the genetic determinant for the cCF10 peptide pheromone in Enterococcus faecalis OG1RF. *Journal of Bacteriology* **184**: 1155-1162.

Arnvig, K.B., and Young, D.B. (2009) Identification of small RNAs in Mycobacterium tuberculosis. *Mol. Micro.* **73**: 397-408.

Bae, T., Clerc-Bardin, S., and Dunny, G.M. (2000) Analysis of expression of prgX, a key negative regulator of the transfer of the Enterococcus faecalis pheromone-inducible plasmid pCF10. *Journal of Molecular Biology* **297**: 861-875.

Bae, T., and Dunny, G.M. (2001) Dominant-negative mutants of prgX: evidence for a role for PrgX dimerization in negative regulation of pheromone-inducible conjugation. *Molecular Microbiology* **39**: 1307-1320.

Bae, T., Kozlowicz, B., and Dunny, G.M. (2002) Two targets in pCF10 DNA for PrgX binding: Their role in production of Qa and prgX mRNA and in regulation of pheromone-inducible conjugation. *Journal of Molecular Biology* **315**: 995-1007.

Bae, T., Kozlowicz, B.K., and Dunny, G.M. (2004) Characterization of cis-acting prgQ mutants: evidence for two distinct repression mechanisms by Qa RNA and PrgX protein in pheromone-inducible enterococcal plasmid pCF10. *Molecular Microbiology* **51**: 271-281.

Bagci, E.Z., Vodovotz, Y., Billiar, T.R., Ermentrout, G.B., and Bahar, I. (2006) Bistability in Apoptosis: Roles of Bax, Bcl-2, and Mitochondrial Permeability Transition Pores. *Biophys. J.* **90**: 1546-1559.

Bagowski, C.P., Besser, J., Frey, C.R., and Ferrell Jr, J.E. (2003) The JNK Cascade as a Biochemical Switch in Mammalian Cells Ultrasensitive and All-or-None Responses *Current Biology* **13**: 315-320.

Becskei, A., Séraphin, B., and Serrano, L. (2001) Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. *EMBO Journal* **20**: 2528-2535.

Beisel, C.L., and Storz, G. (2010) Base pairing small RNAs and their roles in global regulatory networks. *Fems Microbiology Reviews* **34**: 866-882.

Bensing, B., A., Manias, D., A., and Dunny, G., M. (1997) Pheromone cCF10 and plasmid pCF10-encoded regulatory molecules act post-transcriptionally to activate expression of downstream conjugation functions. *Molecular Microbiology* **24**: 285-294.

Bensing, B.A., and Dunny, G.M. (1993) Cloning and molecular analysis of genes affecting expression of binding substance, the recipient-encoded receptor(s) mediating mating aggregate formation in enterococcus-faecalis. *Journal of Bacteriology* **175**: 7421-7429.

Bensing, B.A., Meyer, B.J., and Dunny, G.M. (1996) Sensitive detection of bacterial transcription initiation sites and differentiation from RNA processing sites in the pheromone-induced plasmid transfer system of Enterococcus faecalis. *Proc Natl Acad Sci USA* **93**: 7794-7799.

Bergstrom, C.T., Lipsitch, M., and Levin, B.R. (2000) Natural selection, infectious transfer and the existence conditions for bacterial plasmids. *Genetics* **155**: 1505-1519.

Bernstein, J.A., Khodursky, A.B., Lin, P.H., Lin-Chao, S., and Cohen, S.N. (2002) Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. *Proc Natl Acad Sci U S A* **99**: 9697-9702.

Boles, B., and Horswill, A. (2008) *agr*-Mediated dispersal of *Staphylococcus aureus* biofilms. *PLoS Pathogens* **4**.

Brantl, S. (2002) Antisense-RNA regulation and RNA interference. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* **1575**: 15-25.

Brantl, S. (2007) Regulatory mechanisms employed by cis-encoded antisense RNAs. *Cell regulation (RNA special issue)* **10**: 102-109.

Bryan, E.M., Bae, T., Kleerebezem, H., and Dunny, G.M. (2000) Improved vectors for nisin-controlled expression in gram-positive bacteria. *Plasmid* **44**: 183-190.

Buttaro, B.A., Antiporta, M.H., and Dunny, G.M. (2000) Cell-associated pheromone peptide (cCF10) production and pheromone inhibition in Enterococcus faecalis. *Journal of Bacteriology* **182**: 4926-4933.

Callen, B.P., Shearwin, K.E., and Egan, J.B. (2004) Transcriptional Interference between Convergent Promoters Caused by Elongation over the Promoter. *Molecular Cell* **14**: 647-656.

Carniol, K., and Gilmore, M.S. (2004) Signal transduction, quorum-sensing, and extracellular protease activity in Enterococcus faecalis biofilm formation. *Journal of Bacteriology* **186**: 8161-8163.

Chandler, J.R., and Dunny, G.M. (2004) Enterococcal peptide sex pheromones: synthesis and control of biological activity. *Peptides* **25**: 1377-1388.

Chandler, J.R., Flynn, A.R., Bryan, E.A., and Dunny, G.M. (2005) Specific control of endogenous cCF10 pheromone by a conserved domain of the pCF10-encoded regulatory protein PrgY in Enterococcus faecalis. *Journal of Bacteriology* **187**: 4830-4843.

Chandler, J.R., and Dunny, G.M. (2008) Characterization of the sequence specificity determinants required for processing and control of sex pheromone by the intramembrane protease Eep and the plasmid-encoded protein PrgY. *Journal of Bacteriology* **190**: 1172-1183.

Chater, K.F., and Horinouchi, S. (2003) Signalling early developmental events in two highly diverged Streptomyces species. *Molecular Microbiology* **48**: 9-15.

Chatterjee, A., Kaznessis, Y.N., and Hu, W.S. (2008) Tweaking biological switches through a better understanding of bistability behavior. *Current Opinion in Biotechnology* **19**: 475-481.

Chatterjee, A., Johnson, C.M., Shu, C.-C., Kaznessis, Y.N., Ramkrishna, D., Dunny, G.M., and Hu, W.-S. (2011) Convergent transcription confers a bistable switch in Enterococcus faecalis conjugation. *Proc Natl Acad Sci USA* **108**: 9721-9726.

Chatterjee, A., Laurie, D., Sarika, M., Takano, E., Kaznessis, Y.N., and Hu, W.-S. (2011) Convergent transcription in the Butyrolactone regulon in Streptomyces coelicolor confers a bistable genetic switch for antibiotic biosynthesis. *PLoS ONE* 6: e21974.

Chatterjee, A., Dunny, G.M., and Hu, W.-S. (to be submitted) Quorum sensing via inhibitor signaling molecule allows regulation of pCF10 conjugation in Enterococcus faecalis: a dynamic response to induction.

Chen, I., and Dubnau, D. (2004) DNA uptake during bacterial transformation. *Nature Reviews Microbiology* **2**: 241-249.

Chen, J., Darst, S.A., and Thirumalai, D. (2010) Promoter melting triggered by bacterial RNA polymerase occurs in three steps. *Proc Natl Acad Sci USA* **107**: 12523-12528.

Chen, J.J., Sun, M., Hurst, L.D., Carmichael, G.G., and Rowley, J.D. (2005) Genome-wide analysis of coordinate expression and evolution of human cis-encoded sense-antisense transcripts. *Trends in Genetics* **21**: 326-329.

Chen, Y.Q., Staddon, J.H., and Dunny, G.M. (2007) Specificity determinants of conjugative DNA processing in the Enterococcus faecalis plasmid pCF10 and the Lactococcus lactis plasmid pRS01. *Molecular Microbiology* **63**: 1549-1564.

Chen, Y.Q., Zhang, X.L., Manias, D., Yeo, H.J., Dunny, G.M., and Christie, P.J. (2008) Enterococcus faecalis PcfC, a spatially localized substrate receptor for type IV secretion of the pCF10 transfer intermediate. *Journal of Bacteriology* **190**: 3632-3645.

Chickarmane, V., Troein, C., Nuber, U.A., Sauro, H.M., and Peterson, C. (2006) Transcriptional dynamics of the embryonic stem cell switch. *Plos Computational Biology* **2**: 1080-1092.

Chuang, O.N., Schlievert, P.M., Wells, C.L., Manias, D.A., Tripp, T.J., and Dunny, G.M. (2009) Multiple Functional Domains of Enterococcus faecalis Aggregation Substance Asc10 Contribute to Endocarditis Virulence. *Infection and Immunity* **77**: 539-548.

Chung, J.W., Bensing, B.A., and Dunny, G.M. (1995) Genetic-Analysis of a Region of the Enterococcus-faecalis plasmid pCF10 involved in positive regulation of conjugative transfer-functions. *Journal of Bacteriology* **177**: 2107-2117.

Chung, J.W.W., and Dunny, G.M. (1992) Cis-acting, orientation-dependent, positive control-system activates pheromone-inducible conjugation functions at distances greater than 10 kilobases upstream from its target in enterococcus-faecalis. *Proc Natl Acad Sci USA* **89**: 9020-9024.

Clewell, D.B., An, F.Y., Flannagan, S.E., Antiporta, M., and Dunny, G.M. (2000) Enterococcal sex pheromone precursors are part of signal sequences for surface lipoproteins. *Molecular Microbiology* **35**: 246-247.

Clewell, D.B., Francia, M.V., Flannagan, S.E., and An, F.Y. (2002) Enterococcal plasmid transfer: sex pheromones, transfer origins, relaxases, and the Staphylococcus aureus issue. *Plasmid* **48**: 193-201.

Coiras, M., Lopez-Huertas, M.R., Perez-Olmeda, M., and Alcami, J. (2009) Understanding HIV-1 latency provides clues for the eradication of long-term reservoirs. *Nature Reviews Microbiology* **7**: 798-812.

Cook, L., Chatterjee, A., Barnes, A., Yarwood, J., Hu, W.-S., and Dunny, G. (2011) Biofilm growth alters regulation of conjugation by a bacterial pheromone. *Molecular Microbiology (under review).*

Costerton, J.W., Stewart, P.S., and Greenberg, E.P. (1999) Bacterial biofilms: A common cause of persistent infections. *Science* **284**: 1318-1322.

Cox, R.A. (2004) Quantitative relationships for specific growth rates and macromolecular compositions of Mycobacterium tuberculosis, Streptomyces coelicolor A3(2) and Escherichia coli B/r: an integrative theoretical approach. *Microbiology* **150**: 1413-1426.

Crampton, N., Bonass, W.A., Kirkham, J., Rivetti, C., and Thomson, N.H. (2006) Collision events between RNA polymerases in convergent transcription studied by atomic force microscopy. *Nucl. Acids Res.* **34**: 5416-5425.

D'Alia, D., Nieselt, K., Steigele, S., Muller, J., Verburg, I., and Takano, E. (2009) Noncoding RNA of Glutamine Synthetase I Modulates Antibiotic Production in Streptomyces coelicolor A3(2). *Journal of Bacteriology* **192**: 1160-1164.

David, L., Huber, W., Granovskaia, M., Toedling, J., Palm, C.J., Bofkin, L., Jones, T., Davis, R.W., and Steinmetz, L.M. (2006) A high-resolution map of transcription in the yeast genome. *Proc Natl Acad Sci USA* **103**: 5320-5325.

Ding, Y., and Lawrence, C.E. (2003) A statistical sampling algorithm for RNA secondary structure prediction. *Nucleic Acids Research* **31**: 7280-7301.

Dinger, M.E., Amaral, P.P., Mercer, T.R., Pang, K.C., Bruce, S.J., Gardiner, B.B., Askarian-Amiri, M.E., Ru, K., Solda, G., Simons, C., Sunkin, S.M., Crowe, M.L., Grimmond, S.M., Perkins, A.C., and Mattick, J.S. (2008) Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Research* **18**: 1433-1445.

Dodd, I.B., Shearwin, K.B., and Sneppen, K. (2007) Modelling transcriptional interference and DNA looping in gene regulation. *Journal of Molecular Biology* **369**: 1200-1213.

Dornenburg, J.E., Devita, A.M., Palumbo, M.J., and Wade, J.T. (2010) Widespread Antisense Transcription in Escherichia coli. *MBio* **1**.

Dubnau, D. (1991) Genetic competence in Bacillus-subtilis. *Microbiological Reviews* **55**: 395-424.

Dunny, G., Funk, C., and Adsit, J. (1981) Direct stimulation of the transfer of antibiotic-resistance by sex-pheromones in Streptococcus-faecalis. *Plasmid* **6**: 270-278.

Dunny, G., Yuhasz, M., and Ehrenfeld, E. (1982) Genetic and physiological analysis of conjugation in Streptococcus-faecalis. *Journal of Bacteriology* **151**: 855-859.

Dunny, G.M., Brown, B.L., and Clewell, D.B. (1978) Induced Cell-Aggregation and Mating in Streptococcus-Faecalis - Evidence for a Bacterial Sex-Pheromone. *Proc Natl Acad Sci USA* **75**: 3479-3483.

Dunny, G.M., Zimmerman, D.L., and Tortorello, M.L. (1985) Induction of surface exclusion (entry exclusion) by Streptococcus-faecalis sex-pheromones - use of monoclonal-antibodies to identify an inducible surface-antigen involved in the exclusion process. *Proc Natl Acad Sci USA* **82**: 8582-8586.

Dunny, G.M., and Leonard, B.A.B. (1997) Cell-cell communication in gram-positive bacteria. *Annual Review of Microbiology* **51**: 527-564.

Dunny, G.M., Antiporta, M.H., and Hirt, H. (2001) Peptide pheromone-induced transfer of plasmid pCF10 in Enterococcus faecalis: probing the genetic and molecular basis for specificity of the pheromone response. *Peptides* **22**: 1529-1539.

Dunny, G.M. (2007) The peptide pheromone-inducible conjugation system of Enterococcus faecalis plasmid pCF10: cell-cell signalling, gene transfer, complexity and evolution. *Philosophical Transactions of the Royal Society B-Biological Sciences* **362**: 1185-1193.

Dunny, G.M., and Johnson, C.M. (2011) Regulatory circuits controlling enterococcal conjugation: lessons for functional genomics. *Cell regulation (RNA special issue)* **14**: 174-180.

Eguchi, Y., Itoh, T., and J, T. (1991) Antisense RNA. *Annual Reviews in Biochemistry* **60**: 631-652.

Elowitz, M.B., and Leibler, S. (2000) A synthetic oscillatory network of transcriptional regulators. *Nature* **403**: 335-338.

Eszterhas, S.K., Bouhassira, E.E., Martin, D.I.K., and Fiering, S. (2002) Transcriptional Interference by Independently Regulated Genes Occurs in Any Relative Arrangement of the Genes and Is Influenced by Chromosomal Integration Position. *Mol. Cell. Biol. R 10.1128/MCB.22.2.469-479.2002* **22**: 469-479.

Fixen, K.R., Chandler, J.R., Le, T., Kozlowicz, B.K., Manias, D.A., and Dunny, G.M. (2007) Analysis of the amino acid sequence specificity determinants of the enterococcal cCF10 sex pheromone in interactions with the pheromone-sensing machinery. *Journal of Bacteriology* **189**: 1399-1406.

Franch, T., Petersen, M., Wagner, E.G.H., Jaconsen, J.P., and Gerdes, k. (1999) Antisense RNA regulation in prokaryotes: rapid RNA/RNA interaction facilitated by a general U-turn loop structure. *Journal of Molecular Biology* **294**: 1115-1125.

Fuqua, C., Winans, S.C., and Greenberg, E.P. (1996) Census and consensus in bacterial ecosystems: The LuxR-LuxI family of quorum-sensing transcriptional regulators. *Annual Review of Microbiology* **50**: 727-751.

Fuqua, W.C., Winans, S.C., and Greenberg, E.P. (1994) Quorum sensing in bacteria - the LuxR-LuxI family of cell density-responsive transcriptional regulators. *Journal of Bacteriology* **176**: 269-275.

Gardner, T.S., Cantor, C.R., and Collins, J.J. (2000) Construction of a genetic toggle switch in Escherichia coli. *Nature* **403**: 339-342.

Georg, J., Voss, B., Scholz, I., Mitschke, J., Wilde, A., and Hess, W.R. (2009) Evidence for a major role of antisense RNAs in cyanobacterial gene regulation. *Molecular Systems Biology* **5**.

Ghigo, J.M. (2001) Natural conjugative plasmids induce bacterial biofilm development. *Nature* **412**: 442-445.

Gogarten, J.P., and Townsend, J.P. (2005) Horizontal gene transfer, genome innovation and evolution. *Nature Reviews Microbiology* **3**: 679-687.

Goryachev, A.B., Toh, D.J., Wee, K.B., Lee, T., Zhang, H.B., and Zhang, L.H. (2005) Transition to quorum sensing in an agrobacterium population: a stochastic model. *PLoS Comput Biol* **1**: e37.

Gottelt, M., Kol, S., Gomez-Escribano, J.P., Bibb, M., and Takano, E. Deletion of a regulatory gene within the cpk gene cluster reveals novel antibacterial activity in Streptomyces coelicolor A3(2). *Microbiology-Sgm* **156**: 2343-2353.

Greber, D., and Fussenegger, M. (2007) Mammalian synthetic biology: Engineering of sophisticated gene networks. *Journal of Biotechnology* **130**: 329-345.

Greger, I.H., Aranda, A., and Proudfoot, N. (2000) Balancing transcriptional interference and initiation on the GAL7 promoter of Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U. S. A.* **97**: 8415-8420.

Grohmann, E., Muth, G., and Espinosa, M. (2003) Conjugative plasmid transfer in gram-positive bacteria. *Microbiology and Molecular Biology Reviews* **67**: 277-+.

Guell, M., van Noort, V., Yus, E., Chen, W.H., Leigh-Bell, J., Michalodimitrakis, K., Yamada, T., Arumugam, M., Doerks, T., Kuhner, S., Rode, M., Suyama, M., Schmidt, S., Gavin, A.C., Bork, P., and Serrano, L. (2009) Transcriptome Complexity in a Genome-Reduced Bacterium. *Science* **326**: 1268-1271.

Guet, C., abreve, lin, C., Elowitz, M.B., Hsing, W., and Leibler, S. (2002) Combinatorial Synthesis of Genetic Networks. *Science* **296**: 1466-1470.

Gullerova, M., and Proudfoot, N.J. (2008) Cohesin Complex Promotes Transcriptional Termination between Convergent Genes in S.pombe. *Cell* **132**: 983-995.

Harbottle, H., Thakur, S., Zhao, S., and White, D.G. (2006) Genetics of antimicrobial resistance. *Animal Biotechnology* **17**: 111-124.

Hasty, J., McMillen, D., and Collins, J.J. (2002) Engineered gene circuits. *Nature* **420**: 224-230.

Hernandez, J.A., Muro-Pastor, A.M., Flores, E., Bes, M.T., Peleato, M.L., and Fillat, M.F. (2006) Identification of a furA cis antisense RNA in the cyanobacterium Anabaena sp PCC 7120. *Journal of Molecular Biology* **355**: 325-334.

Hirt, H., Schlievert, P.M., and Dunny, G.M. (2002) In vivo induction of virulence and antibiotic resistance transfer in Enterococcus faecalis mediated by the sex pheromone-sensing system of pCF10. *Infection and Immunity* **70**: 716-723.

Hirt, H., Manias, D., Bryan, E., Klein, J., Marklund, J., Staddon, J., Paustian, M., Kapur, V., and Dunny, G. (2005) Characterization of the Pheromone Response of the *Enterococcus faecalis* Conjugative Plasmid pCF10: Complete Sequence and Comparative Analysis of the Transcriptional and Phenotypic Responses of pCF10-Containing Cells to Pheromone Induction. *Journal of Bacteriology* **187**: 1044-1054.

Hongay, C.F., Grisafi, P.L., Galitski, T., and Fink, G.R. (2006) Antisense Transcription Controls Cell Fate in Saccharomyces cerevisiae. *Cell* **127**: 735-745.

Horinouchi, S. (2002) A microbial hormone, A-factor, as a master switch for morphological differentiation and secondary metabolism in Streptomyces griseus. *Front Biosci* **7**: d2045-2057.

Hsiao, N.H., Soding, J., Linke, D., Lange, C., Hertweck, C., Wohlleben, W., and Takano, E. (2007) ScbA from Streptomyces coelicolor A3(2) has homology to fatty acid synthases and is able to synthesize gamma-butyrolactones. *Microbiology* **153**: 1394-1404.

Hsiao, N.H., Nakayama, S., Merlo, M.E., de Vries, M., Bunet, R., Kitani, S., Nihira, T., and Takano, E. (2009) Analysis of Two Additional Signaling Molecules in Streptomyces coelicolor and the Development of a Butyrolactone-Specific Reporter System. *Chemistry & Biology* **16**: 951-960.

Hughes, D.T., and Sperandio, V. (2008) Inter-kingdom signalling: communication between bacteria and their hosts. *Nat Rev Microbiol* **6**: 111-120.

Jain, R., Rivera, M.C., and Lake, J.A. (1999) Horizontal gene transfer among genomes: The complexity hypothesis. *Proceedings of the National Academy of Sciences of the United States of America* **96**: 3801-3806.

Jayapal, K.P., Philp, R.J., Kok, Y.J., Yap, M.G.S., Sherman, D.H., Griffin, T.J., and Hu, W.S. (2008) Uncovering Genes with Divergent mRNA-Protein Dynamics in Streptomyces coelicolor. *Plos One* **3**.

Jayapal, K.P., Sui, S., Philp, R.J., Kok, Y.J., Yap, M.G.S., Griffin, T.J., and Hu, W.S. (2010) Multitagging Proteomic Strategy to Estimate Protein Turnover Rates in Dynamic Systems. *Journal of Proteome Research* **9**: 2087-2097.

Johnson, C.M., Manias, D.A., Haemig, H.A.H., Shokeen, S., Weaver, K.E., Henkin, T.M., and Dunny, G.M. (2010) Direct Evidence for Control of the Pheromone-Inducible prgQ Operon of Enterococcus faecalis Plasmid pCF10 by a

Countertranscript-Driven Attenuation Mechanism. *Journal of Bacteriology* **192**: 1634-1642.

Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C.C., Suzuki, M., Kawai, J., Suzuki, H., Carninci, P., Hayashizaki, Y., Wells, C., Frith, M., Ravasi, T., Pang, K.C., Hallinan, J., Mattick, J., Hume, D.A., Lipovich, L., Batalov, S., Engstrom, P.G., Mizuno, Y., Faghihi, M.A., Sandelin, A., Chalk, A.M., Mottagui-Tabar, S., Liang, Z., Lenhard, B., and Wahlestedt, C. (2005) Antisense transcription in the mammalian transcriptome. *Science* **309**: 1564-1566.

Kato, J.Y., Funa, N., Watanabe, H., Ohnishi, Y., and Horinouchi, S. (2007) Biosynthesis of gamma-butyrolactone autoregulators that switch on secondary metabolism and morphological development in Streptomyces. *Proc Natl Acad Sci USA* **104**: 2378-2383.

Kim, H.S., Lee, Y.J., Lee, C.K., Choi, S.U., Yeo, S.H., Hwang, Y.I., Yu, T.S., Kinoshita, H., and Nihira, T. (2004) Cloning and characterization of a gene encoding the gamma-butyrolactone autoregulator receptor from Streptomyces clavuligerus. *Arch Microbiol* **182**: 44-50.

Kimura, M., and Yamaguchi, I. (1998) Convergent transcription units and their promoters at both ends of Pot2, an inverted repeat transposon from the rice blast fungus. *Journal of Biochemistry* **124**: 268-273.

Kinoshita, H., Ipposhi, H., Okamoto, S., Nakano, H., Nihira, T., and Yamada, Y. (1997) Butyrolactone autoregulator receptor protein (BarA) as a transcriptional regulator in Streptomyces virginiae. *J Bacteriol* **179**: 6986-6993.

Kitani, S., Yamada, Y., and Nihira, T. (2001) Gene replacement analysis of the butyrolactone autoregulator receptor (FarA) reveals that FarA acts as a Novel regulator in secondary metabolism of Streptomyces lavendulae FRI-5. *J Bacteriol* **183**: 4357-4363.

Kiyosawa, H., Mise, N., Iwase, S., Hayashizaki, Y., and Abe, K. (2005) Disclosing hidden transcripts: Mouse natural sense-antisense transcripts tend to be poly(A) negative and nuclear localized. *Genome Research* **15**: 463-474.

Kleerebezem, M., Quadri, L.E.N., Kuipers, O.P., and deVos, W.M. (1997) Quorum sensing by peptide pheromones and two-component signal-transduction systems in Gram-positive bacteria. *Molecular Microbiology* **24**: 895-904.

Kobayashi, H., Kaern, M., Araki, M., Chung, K., Gardner, T.S., Cantor, C.R., and Collins, J.J. (2004) Programmable cells: Interfacing natural and engineered gene networks. *Proc Natl Acad Sci USA* **101**: 8414-8419.

Koch, S., Hufnagel, M., Theilacker, C., and Huebner, J. (2004) Enterococcal infections: host response, therapeutic, and prophylactic possibilities. *Vaccine* **22**: 822-830.

Koirala, K.D., and Thapa, S.D. (2011) Antibiotic Resistance and the Future. *Jama-Journal of the American Medical Association* **305**: 2293-2293.

Kozlowicz, B.K., Bae, T., and Dunny, G.M. (2004) Enterococcus faecalis pheromone-responsive protein PrgX: genetic separation of positive autoregulatory functions

from those involved in negative regulation of conjugative plasmid transfer. *Molecular Microbiology* **54**: 520-532.

Kozlowicz, B.K. (2005) The molecular mechanism and peptide signaling response of PrgX used to control pheromone-induced conjugative transfer of pCF10. PhD dissertation. In *Department of Microbiology* Minneapolis: University of Minnesota.

Kramer, B.P., Viretta, A.U., Daoud-El Baba, M., Aubel, D., Weber, W., and Fussenegger, M. (2004) An engineered epigenetic transgene switch in mammalian cells. *Nature Biotechnology* **22**: 867-870.

Kramer, B.P., and Fussenegger, M. (2005) Hysteresis in a synthetic mammalian gene network. *Proceedings of the National Academy of Sciences* **102**: 9517-9522.

Krinke, L., and Wulff, D.L. (1987) Oop Rna, Produced from Multicopy Plasmids, Inhibits Lambda-Cii Gene-Expression through an Rnase-Iii-Dependent Mechanism. *Genes & Development* **1**: 1005-1013.

Kristich, C.J., Chandler, J.R., and Dunny, G.M. (2007) Development of a host-genotype-independent counterselectable marker and a high-frequency conjugative delivery system and their use in genetic analysis of Enterococcus faecalis. *Plasmid* **57**: 131-144.

Kurland, C.G., Canback, B., and Berg, O.G. (2003) Horizontal gene transfer: A critical view. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 9658-9662.

Lai, K., Robertson, M.J., and Schaffer, D.V. (2004) The Sonic Hedgehog Signaling System as a Bistable Genetic Switch. *Biophys. J.* **86**: 2748-2757

A theoretical analysis of the bistability behavior and exploration of the role of feedback loop. Introduced stofchasticity in bistability analysis. .

Laslo, P., Spooner, C.J., Warmflash, A., Lancki, D.W., Lee, H.-J., Sciammas, R., Gantner, B.N., Dinner, A.R., and Singh, H. (2006) Multilineage Transcriptional Priming and Determination of Alternate Hematopoietic Cell Fates. *Cell* **126**: 755-766.

Lederberg, J., and Tatum, E.L. (1946) Gene recombination in E. coli. *Nature* **158**: 558.

Lee, E.J., and Groisman, E.A. (2010) An antisense RNA that governs the expression kinetics of a multifunctional virulence gene. *Molecular Microbiology* **76**: 1020-1033.

Lemme, A., Grobe, L., Reck, M., Tomasch, J., and Wagner-Dobler, I. (2011) Subpopulation specific transcriptome analysis of CSP induced *Streptococcus mutans*. *Journal of Bacteriology* **epub**.

Lenasi, T., Contreras, X., and Peterlin, B.M. (2008) Transcriptional interference antagonizes proviral gene expression to promote HIV latency. *Cell Host & Microbe* **4**: 123-133.

Leonard, B.A.B., Podbielski, A., Hedberg, P.J., and Dunny, G.M. (1996) Enterococcus faecalis pheromone binding protein, PrgZ, recruits a chromosomal oligopeptide permease system to import sex pheromone cCF10 for induction of conjugation. *Proc Natl Acad Sci USA* **93**: 260-264.

Liang, S.T., Bipatnath, M., Xu, Y.C., Chen, S.L., Dennis, P., Ehrenberg, M., and Bremer, H. (1999) Activities of constitutive promoters in Escherichia coli. *Journal of Molecular Biology* **292**: 19-37.

Liao, S.M., Wu, T.H., Chiang, C.H., Susskind, M.M., and McClure, W.R. (1987) Control of Gene-Expression in Bacteriophage P22 by a Small Antisense Rna .1. Characterization Invitro of the Psar Promoter and the Sar Rna Transcript. *Genes & Development* **1**: 197-203.

Lujan, S., Guogas, L., Ragonese, H., Matson, S., and Redinbo, M. (2007) Disrupting antibiotic resistance propagation by inhibiting the conjugative DNA relaxase. *Proc Natl Acad Sci USA s* **104**: 12282-12287.

Maamar, H., and Dubnau, D. (2005) Bistability in the Bacillus subtilis K-state (competence) system requires a positive feedback loop. *Mol Microbiol* **56**: 615-624.

McCormick, J.K., Tripp, T.J., Dunny, G.M., and Schlievert, P.M. (2002) Formation of vegetations during infective endocarditis excludes binding of bacterial-specific host antibodies to Enterococcus faecalis. *Journal of Infectious Diseases* **185**: 994-997.

Mehra, S., Charaniya, S., Takano, E., and Hu, W.-S. (2008) A Bistable Gene Switch for Antibiotic Biosynthesis: The Butryolactone Regulon in Streptomyces Coelicolor *Plos One* **3**: 12.

Miller, M.B., and Bassler, B.L. (2001) Quorum sensing in bacteria. *Annual Review of Microbiology* **55**: 165-199.

Misra, S., Crosby, M., Mungall, C., Matthews, B., Campbell, K., Hradecky, P., Huang, Y., Kaminker, J., Millburn, G., Prochnik, S., Smith, C., Tupy, J., Whitfield, E., Bayraktaroglu, L., Berman, B., Bettencourt, B., Celniker, S., de Grey, A., Drysdale, R., Harris, N., Richter, J., Russo, S., Schroeder, A., Shu, S., Stapleton, M., Yamada, C., Ashburner, M., Gelbart, W., Rubin, G., and Lewis, S. (2002) Annotation of the Drosophila melanogaster euchromatic genome: a systematic review. *Genome Biology* **3**: research0083.0081 - 0083.0022.

Misra, S., Crosby, M.A., Mungall, C.J., Matthews, B.B., Campbell, K.S., Hradecky, P., Huang, Y., Kaminker, J.S., Millburn, G.H., Prochnik, S.E., Smith, C.D., Tupy, J.L., Whitfied, E.J., Bayraktaroglu, L., Berman, B.P., Bettencourt, B.R., Celniker, S.E., de Grey, A.D.N.J., Drysdale, R.A., Harris, N.L., Richter, J., Russo, S., Schroeder, A.J., Shu, S.Q., Stapleton, M., Yamada, C., Ashburner, M., Gelbart, W.M., Rubin, G.M., and Lewis, S.E. (2002) Annotation of the Drosophila melanogaster euchromatic genome: a systematic review. *Genome Biol* **3**.

Morfeldt, E., Taylor, D., Vongabain, A., and Arvidson, S. (1995) Activation of alpha-toxin translation in staphylococcus-aureus by the trans-encoded antisense rna, RNAIII. *Embo Journal* **14**: 4569-4577.

Mori, M., Sakagami, Y., Ishii, Y., Isogai, A., Kitada, C., Fujino, M., Adsit, J.C., Dunny, G.M., and Suzuki, A. (1988) Structure of Ccf10, a Peptide Sex-Pheromone Which Induces Conjugative Transfer of the Streptococcus-Faecalis Tetracycline Resistance Plasmid, Pcf10. *Journal of Biological Chemistry* **263**: 14574-14578.

Morohashi, M., Winn, A.E., Borisuk, M.T., Bolouri, H., Doyle, J., and Kitano, H. (2002) Robustness as a measure of plausibility in models of biochemical networks. *Journal of Theoretical Biology* **216**: 19-30.

Nair, V.D., Yuen, T., Olanow, C.W., and Sealfon, S.C. (2004) Early Single Cell Bifurcation of Pro- and Antiapoptotic States during Oxidative Stress. *J. Biol. Chem.* **279**: 27494-27501.

Nakayama, J., Ruhfel, R.E., Dunny, G.M., Isogai, A., and Suzuki, A. (1994) The prgQ gene of the Enterococcus faecalis tetracycline resistance plasmid pCF10 encodes a peptide inhibitor, iCF10. *J. Bacteriol.* **176**: 7405-7408.

Napoli, S., Pastori, C., Magistri, M., Carbone, G.M., and Catapano, C.V. (2009) Promoter-specific transcriptional interference and c-myc gene silencing by siRNAs in human cells. *Embo Journal* **28**: 1708-1719.

Natsume, R., Ohnishi, Y., Senda, T., and Horinouchi, S. (2004) Crystal structure of a gamma-butyrolactone autoregulator receptor protein in Streptomyces coelicolor A3(2). *J Mol Biol* **336**: 409-419.

Neuman, K.C., Abbondanzieri, E.A., Landick, R., Gelles, J., and Block, S.M. (2003) Ubiquitous transcriptional pausing is independent of RNA polymerase backtracking. *Cell* **115**: 437-447.

Nguyen, K., Piastro, K., Gray, T., and Derbyshire, K. (2010) Mycobacterial biofilms facilitate horizontal DNA transfer between strains of *Mycobacterium smegmatis*. *Journal of Bacteriology* **192**: 5134-5142.

Novick, R.P., Ross, H.F., Projan, S.J., Kornblum, J., Kreiswirth, B., and Moghazeh, S. (1993) SYNTHESIS OF STAPHYLOCOCCAL VIRULENCE FACTORS IS CONTROLLED BY A REGULATORY RNA MOLECULE. *Embo Journal* **12**: 3967-3975.

Ochman, H., Lawrence, J.G., and Groisman, E.A. (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**: 299-304.

Olmsted, S.B., Kao, S.M., Vanputte, L.J., Gallo, J.C., and Dunny, G.M. (1991) Role of the pheromone-inducible surface protein-asc10 in mating aggregate formation and conjugal transfer of the Enterococcus-faecalis plasmid-pCF10. *Journal of Bacteriology* **173**: 7665-7672.

Olmsted, S.B., Erlandsen, S.L., Dunny, G.M., and Wells, C.L. (1993) High-resolution visualization by field-emission scanning electron-microscopy of Enterococcus-faecalis surface-proteins encoded by the pheromone-inducible conjugative plasmid pCF10. *Journal of Bacteriology* **175**: 6229-6237.

Onaka, H., Sugiyama, M., and Horinouchi, S. (1997) A mutation at proline-115 in the A-factor receptor protein of Streptomyces griseus abolishes DNA-binding ability but not ligand-binding ability. *J Bacteriol* **179**: 2748-2752.

Ozbudak, E.M., Thattai, M., Lim1, H.N., Shraiman, B.I., and Oudenaarden, A.v. (2004) Multistability in the lactose utilization network of Escherichia coli. *Nature* **427**: 737-740.

Palmer, A.C., Ahlgren-Berg, A., Egan, J.B., Dodd, I.B., and Shearwin, K.E. (2009) Potent Transcriptional Interference by Pausing of RNA Polymerases over a Downstream Promoter. *Molecular Cell* **34**: 545-555.

Panek, J., Bobek, J., Mikulik, K., Basler, M., and Vohradsky, J. (2008) Biocomputational prediction of small non-coding RNAs in Streptomyces. *Bmc Genomics* **9**.

Parsek, M.R., Val, D.L., Hanzelka, B.L., Cronan, J.E., Jr., and Greenberg, E.P. (1999) Acyl homoserine-lactone quorum-sensing signal generation. *Proc Natl Acad Sci U S A* **96**: 4360-4365.

Paulsen, I.T., Banerjei, L., Myers, G.S.A., Nelson, K.E., Seshadri, R., Read, T.D., Fouts, D.E., Eisen, J.A., Gill, S.R., Heidelberg, J.F., Tettelin, H., Dodson, R.J., Umayam, L., Brinkac, L., Beanan, M., Daugherty, S., DeBoy, R.T., Durkin, S., Kolonay, J., Madupu, R., Nelson, W., Vamathevan, J., Tran, B., Upton, J., Hansen, T., Shetty, J., Khouri, H., Utterback, T., Radune, D., Ketchum, K.A., Dougherty, B.A., and Fraser, C.M. (2003) Role of mobile DNA in the evolution of vancomycin-resistant Enterococcus faecalis. *Science* **299**: 2071-2074.

Pomerening, J.R., Sontag, E.D., and Ferrell, J.E. (2003) Building a cell cycle oscillator: hysteresis and bistability in the activation of Cdc2. *Nature Cell Biology* **5**: 346-351.

Prasanth, K.V., and Spector, D.L. (2007) Eukaryotic regulatory RNAs: an answer to the 'genome complexity' conundrum. *Genes & Development* **21**: 11-42.

Prescott, E.M., and Proudfoot, N.J. (2002) Transcriptional collision between convergent genes in budding yeast. *Proc Natl Acad Sci USA* **99**: 8796-8801

Ramage, G., Saville, S.P., Wickes, B.L., and Lopez-Ribot, J.L. (2002) Inhibition of Candida albicans biofilm formation by farnesol, a quorum-sensing molecule. *Applied and Environmental Microbiology* **68**: 5459-5463.

Rasmussen, S., Nielsen, H.B., and Jarmer, H. (2009) The transcriptionally active regions in the genome of Bacillus subtilis. *Molecular Microbiology* **73**: 1043-1057.

CDC report, (2008).

Ruengjitchatchawalya, M., Nihira, T., and Yamada, Y. (1995) Purification and characterization of the IM-2-binding protein from Streptomyces sp. strain FRI-5. *J Bacteriol* **177**: 551-557.

Santillan, M., and Mackey, M.C. (2001) Dynamic regulation of the tryptophan operon: a modeling study and comparison with experimental data. *Proc Natl Acad Sci U S A* **98**: 1364-1369.

Sara, N., Chiara, P., Marco, M., Giuseppina M., C., and Carlo V., C. (2009) Promoter-specific transcriptional interference and c-myc gene silencing by siRNAs in human cells. *EMBO J* **28**: 1708-1719.

Sha, W., Moore, J., Chen, K., Lassaletta, A.D., Yi, C.-S., Tyson, J.J., and Sible, J.C. (2003) From the Cover: Hysteresis drives cell-cycle transitions in Xenopus laevis egg extracts. *Proc Natl Acad Sci USA* **100**: 975-980.

Shearwin, K.E., Callen, B.P., and Egan, J.B. (2005) Transcriptional interference - a crash course. *Trends in Genetics* **21**: 339-345.

Shi, K., Brown, C.K., Gu, Z.Y., Kozlowicz, B.K., Dunny, G.M., Ohlendorf, D.H., and Earhart, C.A. (2005) Structure of peptide sex pheromone receptor PrgX and PrgX/pheromone complexes and regulation of conjugation in Enterococcus faecalis. *Proc Natl Acad Sci USA* **102**: 18596-18601.

Shokeen, S., Johnson, C., Greenfield, T., Manias, D., Dunny, G., and Weaver, K. (2010) Structural analysis of the Anti-Q-Qs interaction: RNA-mediated regulation of *E. faecalis* plasmid pCF10 conjugation. *Plasmid* **64**: 26-35.

Silvaggi, J.M., Perkins, J.B., and Losick, R. (2005) Small untranslated RNA antitoxin in Bacillus subtilis. *Journal of Bacteriology* **187**: 6641-6650.

Simonsen, L. (1991) The existence conditions for bacterial plasmids - theory and reality. *Microbial Ecology* **22**: 187-205.

Sneppen, K., Dodd, I.B., Shearwin, K.E., Palmer, A.C., Schubert, R.A., Callen, B.P., and Egan, J.B. (2005) A mathematical model for transcriptional interference by RNA polymerase traffic in Escherichia coli. *Journal of Molecular Biology* **346**: 399-409.

Sorensen, S.J., Bailey, M., Hansen, L.H., Kroer, N., and Wuertz, S. (2005) Studying plasmid horizontal transfer in situ: A critical review. *Nature Reviews Microbiology* **3**: 700-710.

Sotiropoulos, V., and Kaznessis, Y. (2007) Synthetic tetracycline-inducible regulatory networks: computer-aided design of dynamic phenotypes. *BMC Systems Biology* **1**: 18.

Staddon, J.H., Bryan, E.M., Manias, D.A., and Dunny, G.M. (2004) Conserved target for group II intron insertion in relaxase genes of conjugative elements of gram-positive bacteria. *Journal of Bacteriology* **186**: 2393-2401.

Stelling, J., Sauer, U., Szallasi, Z., Doyle, F.J., and Doyle, J. (2004) Robustness of cellular functions. *Cell* **118**: 675-685.

Stougaard, P., Molin, S., and Nordstrom, K. (1981) Rnas Involved in Copy-Number Control and Incompatibility of Plasmid R1. *Proc Natl Acad Sci USA* **78**: 6008-6012.

Swiercz, J.P., Hindra, Bobek, J., Haiser, H.J., Di Berardo, C., Tjaden, B., and Elliot, M.A. (2008) Small non-coding RNAs in Streptomyces coelicolor. *Nucleic Acids Research* **36**: 7240-7251.

Takano, E., Nihira, T., Hara, Y., Jones, J.J., Gershater, C.J., Yamada, Y., and Bibb, M. (2000) Purification and structural determination of SCB1, a gamma-butyrolactone that elicits antibiotic production in Streptomyces coelicolor A3(2). *J Biol Chem* **275**: 11010-11016.

Takano, E., Chakraburtty, R., Nihira, T., Yamada, Y., and Bibb, M.J. (2001) A complex role for the gamma-butyrolactone SCB1 in regulating antibiotic production in Streptomyces coelicolor A3(2). *Mol Microbiol* **41**: 1015-1028.

Takano, E., Kinoshita, H., Mersinias, V., Bucca, G., Hotchkiss, G., Nihira, T., Smith, C.P., Bibb, M., Wohlleben, W., and Chater, K. (2005) A bacterial hormone (the SCB1) directly controls the expression of a pathway-specific regulatory gene in

the cryptic type I polyketide biosynthetic gene cluster of Streptomyces coelicolor. *Mol Microbiol* **56**: 465-479.

Takano, E. (2006) gamma-Butyrolactones: Streptomyces signalling molecules regulating antibiotic production and differentiation. *Curr Opin Microbiol* **9**: 287-294.

Tang, G.Q., Roy, R., Bandwar, R.P., Ha, T., and Patel, S.S. (2009) Real-time observation of the transition from transcription initiation to elongation of the RNA polymerase. *Proc Natl Acad Sci USA* **106**: 22175-22180.

Tendolkar, P.M., Baghdayan, A.S., and Shankar, N. (2003) Pathogenic enterococci: new developments in the 21st century. *Cellular and Molecular Life Sciences* **60**: 2622-2636.

Thisted, T., and Gerdes, K. (1992) Mechanism of Post-Segregational Killing by the Hok Sok System of Plasmid R1 - Sok Antisense Rna Regulates Hok Gene-Expression Indirectly through the Overlapping Mok Gene. *Journal of Molecular Biology* **223**: 41-54.

Thomason, M.K., and Storz, G. (2010) Bacterial Antisense RNAs: How Many Are There, and What Are They Doing? *Annual Review of Genetics, Vol 44* **44**: 167-188.

Tian, T., and Burrage, K. (2004) Bistability and switching in the lysis/lysogeny genetic regulatory network of bacteriophage lambda. *J Theor Biol* **227**: 229-237.

Toledo-Arana, A., Dussurget, O., Nikitas, G., Sesto, N., Guet-Revillet, H.l.n., Balestrino, D., Loh, E., Gripenland, J., Tiensuu, T., Vaitkevicius, K., Barthelemy, M., Vergassola, M., Nahori, M.-A., Soubigou, G., Re´gnault, B.a., Coppe´e, J.-Y., Lecuit, M., Johansson, J.r., and Cossart, P. (2009) The Listeria transcriptional landscape from saprophytism to virulence. *Nature* **459**: 950-956.

Tomizawa, J.I., and Itoh, T. (1981) Plasmid Cole1 Incompatibility Determined by Interaction of Rna-I with Primer Transcript. *Proc Natl Acad Sci USA* **78**: 6096-6100.

Torres, O.R., Korman, R.Z., Zahler, S.A., and Dunny, G.M. (1991) The conjugative transposon Tn925 - enhancement of conjugal transfer by tetracycline in enterococcus-faecalis and mobilization of chromosomal genes in Bacillus-subtilis and E-faecalis. *Molecular & General Genetics* **225**: 395-400.

Vo, N.V., Hsu, L.M., Kane, C.M., and Chamberlin, M.J. (2003) In vitro studies of transcript initiation by Escherichia coli RNA polymerase. 3. Influences of individual DNA elements within the promoter recognition region on abortive initiation and promoter escape. *Biochemistry* **42**: 3798-3811.

Voigt, C.A., Wolf, D.M., and Arkin, A.P. (2005) The Bacillus subtilis sin operon: an evolvable network motif. *Genetics* **169**: 1187-1202.

Wang, L., Walker, B.L., Iannaccone, S., Bhatt, D., Kennedy, P.J., and Tse, W.T. (2009) Bistable switches control memory and plasticity in cellular differentiation. *Proc Natl Acad Sci USA* **106**: 6638-6643.

Wang, M.D., Schnitzer, M.J., Yin, H., Landick, R., Gelles, J., and Block, S.M. (1998) Force and velocity measured for single molecules of RNA polymerase. *Science* **282**: 902-907.

Wang, X.J., Gaasterland, T., and Chua, N.H. (2005) Genome-wide prediction and identification of cis-natural antisense transcripts in Arabidopsis thaliana. *Genome Biology* **6**.

Ward, D.F., and Murray, N.E. (1979) Convergent transcription in bacteriophage [lambda]: Interference with gene expression. *Journal of Molecular Biology* **133**: 249-266.

Waters, C.M., and Bassler, B.L. (2005) Quorum sensing: Cell-to-cell communication in bacteria. *Annual Review of Cell and Developmental Biology* **21**: 319-346.

Waters, L.S., and Storz, G. (2009) Regulatory RNAs in Bacteria. *Cell* **136**: 615-628.

Watnick, P., and Kolter, R. (2000) Biofilm, city of microbes. *Journal of Bacteriology* **182**: 2675-2679.

Wee, K.B., and Aguda, B.D. (2006) Akt versus p53 in a Network of Oncogenes and Tumor Suppressor Genes Regulating Cell Survival and Death. *Biophys. J.* **91**: 857-865.

Wong, P., Gladney, S., and Keasling, J.D. (1997) Mathematical model of the lac operon: inducer exclusion, catabolite repression, and diauxic growth on glucose and lactose. *Biotechnol Prog* **13**: 132-143.

Xiong, W., and Ferrell Jr., J.E. (2003) A positive-feedback-based bistable 'memory module' that governs a cell fate decision. *Nature* **426**: 460-465.

Yamada, K., Lim, J., Dale, J.M., Chen, H.M., Shinn, P., Palm, C.J., Southwick, A.M., Wu, H.C., Kim, C., Nguyen, M., Pham, P., Cheuk, R., Karlin-Newmann, G., Liu, S.X., Lam, B., Sakano, H., Wu, T., Yu, G.X., Miranda, M., Quach, H.L., Tripp, M., Chang, C.H., Lee, J.M., Toriumi, M., Chan, M.M.H., Tang, C.C., Onodera, C.S., Deng, J.M., Akiyama, K., Ansari, Y., Arakawa, T., Banh, J., Banno, F., Bowser, L., Brooks, S., Carninci, P., Chao, Q.M., Choy, N., Enju, A., Goldsmith, A.D., Gurjal, M., Hansen, N.F., Hayashizaki, Y., Johnson-Hopson, C., Hsuan, V.W., Iida, K., Karnes, M., Khan, S., Koesema, E., Ishida, J., Jiang, P.X., Jones, T., Kawai, J., Kamiya, A., Meyers, C., Nakajima, M., Narusaka, M., Seki, M., Sakurai, T., Satou, M., Tamse, R., Vaysberg, M., Wallender, E.K., Wong, C., Yamamura, Y., Yuan, S.L., Shinozaki, K., Davis, R.W., Theologis, A., and Ecker, J.R. (2003) Empirical analysis of transcriptional activity in the Arabidopsis genome. *Science* **302**: 842-846.

Yelin, R., Dahary, D., Sorek, R., Levanon, E.Y., Goldstein, O., Shoshan, A., Diber, A., Biton, S., Tamir, Y., Khosravi, R., Nemzer, S., Pinner, E., Walach, S., Bernstein, J., Savitsky, K., and Rotman, G. (2003) Widespread occurrence of antisense transcription in the human genome. *Nature Biotechnology* **21,** : 379 - 386.

Young, H.K. (1993) ANTIMICROBIAL RESISTANCE SPREAD IN AQUATIC ENVIRONMENTS. *Journal of Antimicrobial Chemotherapy* **31**: 627-635.