A Perspective on Competence Development in *Streptococcus pneumoniae*: Entrance Versus Exit

BY

LIMING WENG

B.S., Xi'an Jiaotong University, 1997

M.S., Peking Union Medical College, 2006

THESIS

Submitted as partial fulfillment of the requirements for the degree of Doctor of Philosophy in Biological Sciences in the Graduate College of the University at Illinois at Chicago, 2011

Chicago, Illinois

Defense Committee:

       Marion Hulett, Chair
       Donald Morrison, Advisor
       Brian Kay
       Peter Okkema
       Nancy Freitag, Dept. of Microbiology & Immunology

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Dr. Donald Morrison, for his support, guidance and invaluable scientific input in my project. I consider it an immerse privilege to have had the opportunity to learn from such a passionate and accomplished scientist. Dr. Morrison has trained to me to conduct independent research and to get a feel for what it takes to be in charge of broad research projects. He has been my inspiration as I hurdle all the obstacles in the completion of this research work. Furthermore, I am grateful for his encouragement to learn from numerous activities outside the lab, including courses, conferences, and workshops.

I am also deeply indebted to other members of my advisory committee, Dr. Marion Hulett, Dr. Nancy Freitag, Dr. Brian Kay and Dr. Peter Okkema. They each spent a lot of time giving me valuable suggestions and tremendous encouragement. They never ceased to impress me with their professional ways of looking at my project and with practical answers to my numerous requests for help.

I also extend my gratitude to other lab members. Drs. Andrew Piotrowski and Sarita Ahlawat helped me a lot in getting familiar with surroundings when I first joined the lab. They were great sources of ideas and comments in the lab. Haiying Li was very helpful in providing technical support for my experiment, as well as offering honest advice regarding to my career path. Several previous lab members, including Drs. Bhushan Desai, Changkyoo Sung, Ping Luo, Ekaterina Pestova, also deserve my thanks. Even though we never overlapped in the lab, their accomplishments provide much of the basis for my project. I would also like to acknowledge all the LMB members of Department of Biological Science, including David Taussig, Vidhu Mathur, Zhanna Lipatova and many others. In

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVATIONS

A   adenine

a.a.   amino acid(s)

AD   DNA activation domain of Gal4

α-gal   alpha-galactosidase

ATP   adenosine triphosphate

BD   DNA binding domain of Gal4

β-gal   beta-galactosidase

bp   base pair(s)

BSA   bovine serum albumin

C   cytosine

cAMP   cyclic adenosine monophoshpate

CAT   casein hydrolysate yeast extract medium

CFU   colony forming unit

Cm   chloramphenical

CRE   competence regulatory element

CRP   catabolite regulation protein

CSF   competence stimulating factor

CSP   competence stimulating peptide

C-terminal   carboxy terminal

DNA   deoxyribonucleic acid

DNaseI   deoxyribonuclease I

DTT   dithiothreitol

DR   direct repeat

EB/EtBr	ethidium bromide
EDTA	ethylene diamine tetraacetic acid
Em	erythromycin
G	guanine
HK	histidine kinase
HRP	horseradish peroxidase
kb	kilo-base pairs
kDa	kilo-Daltons
Kan	kanamycin
L	liter
LB	Luria-Bertani
Ni-NTA	nickel-nitrilotriacetic acid
Nov	novobiocin
N-terminal	amino terminal
OD	optical density
ONPG	ortho-nitrophenyl-β-D-galactoside
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
Pc	constitutive promoter
PMSF	phenylmethyl sulfonyl fluoride
PNPG	p-nitrophenyl-β-D-galactopyranoside
PTM	post-translational modification
PVDF	polyvinyl diflouride
RBS	ribosome binding site

RNA	ribonucleic acid
RNAP	RNA polymerase
RT	room temperature
SDS	sodium dodecyl sulfate
SOEing	Splicing by Overlap Extension
Spc	spectinomycin
T	thymine
TBE	tris-boric acid-EDTA
TBST	tris buffered saline-Tween
TCTS	two component transduction system
Tet	tetracycline
TE	tris-EDTA
TEMED	N, N, N', N'-tetramethylethylene diamine
TGED	tris-glycerol-EDTA-DTT
Tm	melting temperature of primers
Tris	tris (hydroxymethyl) aminomethane
WT	wild type
YPD	yeast extact-peptone-dextrose
::	novel junction (fusion or insertion)

SUMMARY

In bacteria, genetic transformation is a physiological process of taking up exogenous DNA from the extracellular environment. In *S. pneumoniae*, this process is a transient state and subject to tight regulation. To initiate competence, two sets of gene expressions need to be induced by a signal peptide, CSP. The early gene expression produces ComX and ComW. ComX and ComW work in concert to turn on late gene expression, which eventually make cell competent.

Among late genes, there are many uncharacterized ones whose functions remain unknown. To find out if those unknown late genes are required for competence development, classic genetic analysis is necessary. In order to have a gene disruption tool without leaving the selective marker in the genome, I sought to construct a *lox/cre/lox* cassette for creating efficient marker-less gene deletion. In this new cassette, two mutant lox sites, *lox66* and *lox71*, flank a erythromycin-resistance marker, *ermAM*, that can be used as a temporary marker for selection of desired recombinants, and a *cre* gene, which is under the control of a native regulated pneumococcal promoter. The cassette may subsequently be removed by the recombination of the two lox sites induced by the controllable induction of the *cre* gene, with retention of 34 bp from the cassette as an inert residual double-mutant *lox72* site. We named this cassette "Cheshire". Compared to other gene deletion strategies used in streptococcus, the Cheshire cassette has shown to be more efficient and reliable.

As soon as competence is turned on, it is shut off immediately. Proteolysis of ComX and ComW has been proposed to be one factor that shuts off competence in time. However, it seems that there is another factor from the late gene products that might also play a role in competence termination. I decided to screen on the ComX-induced late genes, especially the transformation-essential late genes.

Among the 20 genes tested, Δ*dprA* displayed a prolonged late gene expression pattern, whereas mutants lacking *cbpD, cibABC, cglEFG, coiA, ssbB, celAB, cclA, cglABCD, cflAB, or radA*, exhibited a wild-type temporal expression pattern. Combined with previous lab work showing that DprA limits the amounts of ComX and another early gene product ComW, I concluded that DprA controls their amounts by inhibiting early gene expression rather than by promoting the degradation of ComX and ComW. To ask what target DprA might work on in turning off early gene expression, yeast two-hybrid assay was employed to investigate protein interactions between DprA and ComD, DprA and ComE. My results suggest that DprA could interact with ComE, but not with ComD, ComX, and ComW. Therefore, I further hypothesize that DprA shuts off early gene expression and competence by a direct interaction with ComE, which makes ComE unable to be phosphorylated.

On the other hand, ComW has been found to be required for competence development along with ComX. However, how ComW participates in this process is not known. To ask if ComX and ComW interact directly, I took advantage of yeast two-hybrid system by fusing ComX with the activation domain and ComW with the DNA-binding domain of the GAL4 transcription factor. My results revealed a direct interaction between ComX and ComW. Taken together with previous results, my data indicate that the role of ComW in turning on competence might be via a direct interaction with ComX. Considering that ComW is a small protein *S. pneumoniae*, I hypothesize that ComW might work as an adaptor for the activation of ComX.

Chapter I: Introduction

1.1 *Streptococcus pneumoniae*

Streptcococcus pneumoniae is a Gram-positive, alpha-hemolytic bacterium. It appears in pairs of short-chains in mango-shaped colonies. It has no spores, no vacuoles, no visible granules, no flagella, and therefore is non-motile (White, 1938). Three surface layers have been observed: plasma membrane, cell wall and capsule. Optochin and bile solubility tests have been used for identification of this species from others (Lennette et al., 1980; Howden, 1979). Its whole genome contains about 2 million base pairs, coding for about 2,200 proteins (Hoskins et al., 2001).

It normally inhabits the upper respiratory tract as an opportunistic pathogen, which means it can cause many kinds of infections when it grows out of control, such as pneumonia, otitis and meningitis. Every year, millions of people died from its infections worldwide. Most of them are children less than five years old and senior citizen greater than six-five years old (Scott, 2007). During these infections, cohabitation and competition of *S. pneumoniae* with other bacteria, such as *Haemophilus influenza*, *Moraxella catarrhalis*, and *Staphyloccocus aureus*, have been observed (Murphy et. al., 2009; Pettigrew et al., 2008).

In *S. pneumoniae*, many virulence factors have been found to contribute to its mobility and mortality. CPS (capsular polysaccharide) was the first factor found to be important for virulence. CPS helps the pneumocccal cell to survive the lung, spread to bloodstream and even contribute to antibiotic tolerance (Morona et al., 2004). It is based on the CPS composition that pneumococcal strains have been divided into 91 serotypes (Park et al., 2007). Ply (pneumolysin) is a cytoplasmic toxin and can cause the formation of pores on the membrane of eukaryotic cells resulting in discharge of the cytoplasm and tissue damage once it is released (Hirst et al., 2004). LytA is a major autolysin that triggers the cell lysis and therefore helps the release of Ply from pneumococcal cells. It is also involved in the prevention of

phagocytosis and production of cytokines (Martinez et al., 2009). PspA and PspC are choline-binding proteins that play roles in colonization by facilitating adherence to epithelial cells and prevent phagocytosis (Ogunniyi et al., 2009). BlpM and BlpN are bacteriocins that are secreted to kill closely related species and enhance intra and interspecies competition in nasophaynx (Dawid et al., 2007).

Penicillin is the first antibiotic that has been used to treat pneumococcal infections since WWII. After that, many types of antibiotics have developed to fight against this species, such as beta-lactams, macrolides, chloramphenicol, tetracyclines and fluroquinolones. Meanwhile, the resistance of pneumococcus to these antibiotics has increased worldwide (EARSS, 2008), due to its easily transformable nature. Pneumococcal infection used to be readily treated with penicillin. But now it's rapidly becoming resistant to many drugs. It is commonly believed that resistance to major class antibiotics is acquired by transformation, taking up foreign DNA from environment and incorporate into its own genome (Dowson, 2004). Unlike many other pathogens with acquired drug resistance, both sensitive and resistant *pneumococci* can be carried asymptomatically in the nasopharynx, which is also the primary resource of person-to-person transmission (Tomasz, 1997). This facilitates the exchange of genetic material between them. And it was found that the majority of pneumococcal resistance is rarely the result of single mutations (Tomasz, 1997). It is more likely to be multiple alterations, or mosaic structures of the gene, which is usually an indicator of genetic exchange. Multiple alterations of penicillin binding proteins (PBPs) are one of the major causes of penicillin resistance in *S. pneumoniae*. It was observed that PBP genes of genetically distinct penicillin-sensitive strains are almost identical in DNA sequence, differing by less than 0.1% in nucleotide sequence. In contrast, PBP genes from penicillin-resistance isolates contain regions that are approximately 20% divergent in their DNA sequence when compared to those of sensitive strains (Hakenbeck, 2000). This also indicated that PBP was acquired by gene transfer, not by the accumulation of random mutations. What's more, PBP

resistant genes are distributed among many other β-lactam resistant strains of commensal streptococcal species, such as *Streptococcus mitis*, *Streptococcus sanguis* and *Streptococcus oralis*. It seemed that there was a PBP gene pool shared by a variety of related streptococci. The commensal streptococcal species are not naturally resistant but susceptible as *pneumococcus*. These suggested that the major genetic way to acquire resistance in *pneumococcus* is through the exchange of large pieces of DNA between closely related strains. Experiments in the laboratory have also proved that antibiotic and especially β-lactam resistance markers can be exchanged between competent streptococcal species. This led to the model of antibiotic resistance development as originated in commensal species (Hakenbeck, 1999). So the correlation between antibiotic resistance and transformation is corroboratory. Meanwhile, the acquired new traits and abilities enable this species to survive various internal and external factors, such as phagocytosis and the "hostile" antibiotic attack, and become diversified.

Currently, there are two types of vaccine available on the market against *S. pneumoniae*: pneumococcal polysaccharide vaccine (such as PPV-23) and pneumococcal conjugate vaccine (such as PCV-13). PPV-23 (Merck, Inc.) is a vaccine mixture containing polysaccharides from 23 serotypes and is effective among them. It is recommended to senior citizen and children over 2 years old (WHO, 2008). PCV-13 (Pfizer, Inc.) is a new vaccine that has polysaccharides from 13 serotypes conjugated to a carrier protein and is thought to be safe to all populations (Duggan, 2010).

1.2 Cell-cell communication in bacteria (Quorum sensing)

In bacteria, cell-cell communication, or quorum sensing, has been universally adopted as a way to sense the chemical signals secreted from other cells in the environment and take synchronal actions in response, similar to exchanging thoughts and taking joint action in human world. It is usually mediated by ligand-recepter signal transduction system. The Gram-negative bacteria usually synthesize a chemical molecule and spread it around, such as the LuxIR system in *Vibrio fischeri* (Bassler, 1999).

When the molecule accumulates to certain concentration, it can bind to a macromolecule, activate the transcriptional factor and trigger multiple gene expressions. In Gram-positive bacteria, a two-component transduction system (TCTS) is usually adopted to transmit the signal inside, such as the Agr quorum sensing system in *Staphylococcus aureus* (Gustafsson et al., 2004). They usually use modified peptides instead of chemicals as the signal molecules. In most cases, a premature peptide, which is synthesized in cytoplasm, needs to be transported out and modified to its active form. This modified peptide then recognizes its specific membrane receptor, which is usually a kinase, and activates a response regulator inside to transcribe a set of genes. The cell-cell communications in bacteria, even though it may not be as sophisticated and elaborated as the multicellular organisms, can guarantee a large population of the cells to respond and take actions simultaneously in many critical aspects of bacteriology, such as bioluminescence, sporulation, biofilm formation, secretion of virulence factors, and competence development (Ng et al., 2009).

1.3 General mechanism of genetic transformation in bacteria

In bacteria, genetic transformation is a physiological process of taking up exogenous DNA from the extracellular environment, which differs it from other genetic exchanges, such as conjugation and transduction. Its original discovery can be dated back in 1928 by Frederick Grifith when he was working with *Streptococcus pneumoniae*. In his experiment, Grifith found a smooth and harmless strain can be transformed by the heat-killed, rough and virulent strain. It was actually this experiment that led to the establishment of DNA as the genetic material, not proteins.

So far, the ability to become competent for genetic transformation has been found in more than 60 species, including both Gram-positive and Gram-negative bacteria (Johnsborg et al., 2007). Well-characterized among them are *Bacillus subtilis*, *Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Neisseria gonorrhoeae* (Lorenz and Wachernagel, 1994). The state of competence can naturally

occur when cells were growing at high density, in nutrition-limited media or other stringent conditions, such as antibiotic treatment (Stewart and Carlson, 1986; Charpentier et al., 2011). It is thought that competence development might be one type of stress response (Claverys et al., 2006). If manipulated properly, competence in some bacteria can be artificially induced by small molecules, such as cAMP or peptides (Mercenier and Chassy, 1988).

Basically, the process of transformation consists of three steps: DNA binding to cell wall, DNA transporting into cytoplasm and DNA integrating into genomic DNA. Each step requires multiple proteins to be functional.

1.3.1 DNA binding

Sources of donor DNA for transformation could be from lysed dead cells, or from secretion by other cells via type IV pili (Dillard and Seifert, 2001). For natural transformation in bacteria, the donor DNA has to be double-stranded, either linear or circular. On the cell membrane, there are receptor proteins that can recognize the dsDNA and bind to it. The DNA sequence for the recognition is unspecific in Gram-positive bacteria. But in Gram-negative system, a specific sequence, called DUS, is usually required for efficient transformation (Sisco and Smith, 1979; Elkins et al., 1991). In *S. pneumoniae*, upon donor DNA binding to the membrane, a nick is created randomly on the dsDNA at about 6 kb in length (Morrison and Guild, 1973). Then an endonuclease, EndA, joins the cutting process, degrading one strand of the dsDNA. The remaining ssDNA is protected somehow and transferred into the cell immediately (Lacks et al., 1975; Puyet et al., 1990). Different species may have different numbers of the binding sites on the membrane. In *S. pneumoniae*, the number was estimated to be around 50 per cell (Fox and Hotchkiss, 1957).

1.3.2 DNA translocation

It has been observed that only ssDNA can pass through the membrane translocasome, either in the direction of 5' to 3' or 3' to 5'. The 3' to 5' direction is what we have observed in the transformation of *S. pneumoniae* (Méjean et al., 1988*),* while some bacteria can do it both ways (Vagner et al., 1990). The ssDNA transfer apparatus is like Type IV secretion pili, composed of multiple protein components with different roles. It is believed that this protein complex is assembled on the membrane. Figure 1 shows the composition of the DNA translocasome in *S. pneumoniae*. Protein complex CglCDEFG, which are transported and processed by CglAB and CilC, make up the out-layer of the translocasome and serve as the scaffold of the pore, while CelA and CelB occupy the inner-layer and interact with incoming DNA. It is also speculated that the CglCDEFG complex might function as the cell wall receptor recruiting donor DNA. EndA is the endonuclease on the cell wall that is supposed to degrade the non-entering strand successively from 5' to 3'. CflA is a helicase which might help to separate DNA strands and driven one strand into the cytoplasm. Function of CflB in transformation is not well known but is proposed to be a pilot protein leading ssDNA (Lacks, 2004).

1.3.3 DNA integration

The incoming ssDNA stand is not stable by itself due to the exposure to the cytoplasmic nucleases; therefore it usually has to be coated and protected by SsbB, DprA and other proteins (Fig. 1). Because this nucleoprotein complex has no transformation activity, it is also termed "eclipse complex". SsbB is widely conserved in Gram-positive bacteria and has been found to be the major component of the eclipse complex (Morrison et al., 2007). DprA, or Smf (originally found in *E coli*), is a DNA processing protein which is believed to facilitate recombination. DprA is able to interact with DNA, SsbB, and RecA. In DprA mutant, ssDNA is not stabilized. DprA may be also a minor component of the
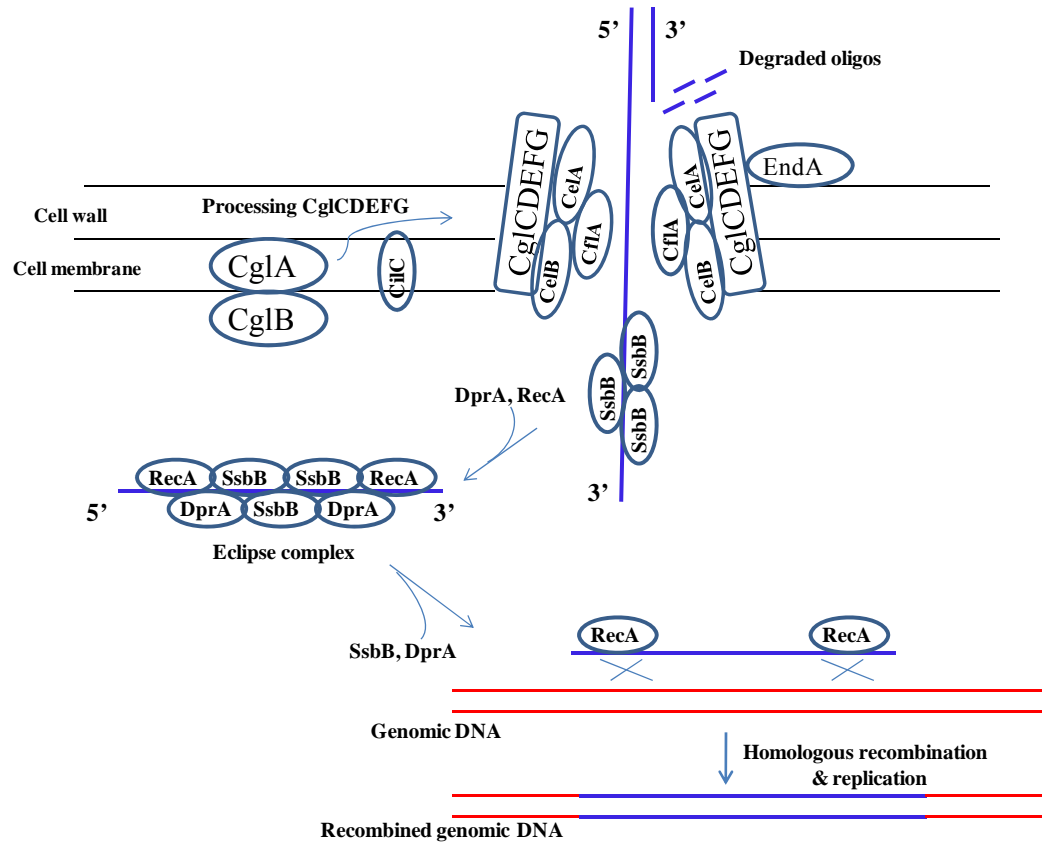
Figure 1 Model of DNA incorporation in *S. pneumoniae*. Upon DNA binding, one strand of the dsDNA is degraded by EndA nuclease from 5' to 3'. Another strand is transferred into cytoplasm from 3' to 5' direction via a hypothetical structure extruding through the cell wall, termed translocasome. This scaffold of the translocasome is mainly composed of CglC-G proteins, which are exported by CglAB and processed by CilC by cleaving the polypeptide at the F residue and methylating the new N-terminus. The incoming ssDNA is immediately bound and protected by SsbB from the degradation of cytoplasmic nucleases. Upon SsbB binding, other two important proteins, DrpA and RecA, were recruited to form the nucleoprotein complex, which is named "Eclipse Complex" because it is recombination-inactive. To make recombination happen, DprA and SsbB need to fall off to expose the ssDNA while keeping RecA to catalyze the homologous crossovers with genomic DNA. The process of homologous recombination is similar to the formation of Holliday junction, detailed in Dressler D. and H. Potter (1982).

eclipse complex, serving as an adaptor gluing everything together. The interaction between DprA and RecA suggests another role DprA might play during transformation: forwarding incoming ssDNA to RecA. RecA is a recombinase and essential for recombination in general. For effective transformation to occur, the protecting proteins on the eclipse complex, such as SsbB and DprA, have to fall off to expose the ssDNA to the recombinase and genomic DNA. It is believed that the existence of DprA in the eclipse complex help alleviate the emerge barrier of the disassembly of the nucleoprotein complex (Mortier-Barrière et al., 2007).

1.4 Gene regulation of competence in bacteria

1.4.1 Competence regulation in *Haemophilus influenzae*

Haemophilus influenzae is a Gram-negative and opportunistic pathogen causing infections similar to *S. pneumoniae*. Competence occurs in nearly all of the cells when the culture grows to high cell density, or grows in nutrient-limited media (Herriott et al., 1970). In this bacterium, competence induction has been shown to be dependent on the cyclic adenosine monophosphate (cAMP) and its receptor protein, CRP (Chandler, 1992). As a secondary messenger in cell, the levels of cAMP rise in response to nutrient stress (Peterkofsky et al., 1971). CRP is a transcription factor. When cAMP binds to CRP, the activated CRP has been found to recognize a conserved palindrome sequence (**c**ompetence **r**egulatory **e**lement, CRE) in front of the promoters of several genes and up-regulates their expressions (Tomb et al., 1993). So far, 25 genes in 13 operons have been identified as components of the competence regulon, involved in DNA uptake, DNA processing and recombination (Macfadyen, 2000). One of them is the *sxy* gene. Sxy has been hypothesized as a transcription regulator that also contains a CRE site. It has been found that its expression increased 3-fold in the presence of cAMP and CRP. Upon expression, Sxy is believed to play a role in assisting CRP recognizing CRE sites (Figure 2), therefore
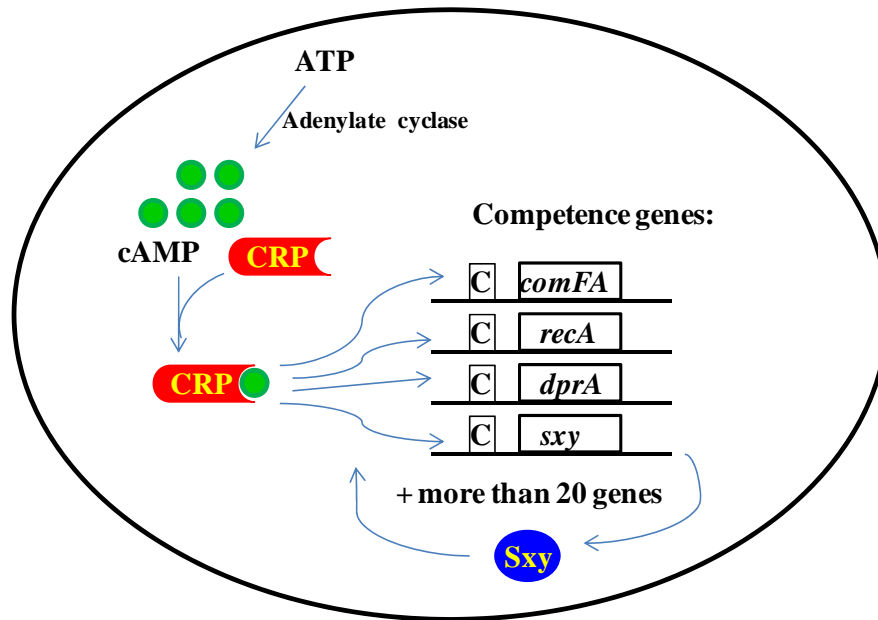
Figure 2 Competence regulation in *Haemophilus influenzae*. Competence in *H. influenzae* starts with the synthesis of the secondary messenger, cAMP, which usually rises in response to nutrient stress in the cytoplasm. Once cAMP is recruited into its receptor protein (CRP, catabolite regulation protein), the activated CRP-cAMP complex is able to recognize a conserved sequence (CRE, competence regulatory element) in front of the promoter of competence genes and up-regulate their expressions. One of competence gene products is Sxy, which has been found able to assist CRP recognizing CRE sites, therefore forming a positive feedback loop in turning on competence. C box, CRE sequence.

forming a positive feedback loop in turning on competence. It has been found that both mutations in CRP and Sxy completely disrupt competence development (Redfield et al., 2005).

1.4.2 Competence regulation in *Neisseria gonorrhoeae*

Neisseria gonorrhoeae is also Gram-negative bacterium and responsible for sexually transmitted infections. A remarkable feature of this species in natural transformation lies in the fact that it is always in a competent state. No special conditions are required for its induction. That means the competence genes in *Neisseria gonorrhoeae* are constitutively expressed and the DNA uptake machine has to be always in place. However, the DNA uptake apparatus and transformation process are not different with those in *S. pneumonae* and *B. subtilis*. The structure, as well as the detailed mechanism, of the type IV pilus-like translocasome, was reviewed in Finn et al., 2002. The donor DNA usually comes from the secretion of related species, including human cells. It is believed that the DNA binding and uptake is sequence specific. A 10-bp consensus sequence which is rich in the genome of *N. gonorrhoeae* has been found necessary for effective transformation. The polarity of the incoming ssDNA is 5' to 3', with 5' capped and protected (Chen and Dubnan, 2004)

1.4.3 Competence regulation in *Bacillus subtilis*

Bacillus subtilis is a Gram-positive bacterium which is most commonly found in soil. It is rod-shaped and has the ability to form endodpores and biofilm in certain conditions. It is not a human pathogen but can cause food contamination. Competence in *B. subtilis* has been observed to occur at the onset of stationary phase of growth in nature, indicating it might be induced by nutrient starvation (Turgay et al., 1998). The DNA translocasome in *B. subtilis* resembles the type IV pilus, similar to the counterpart in *N. gonorrheae* (Hamoen et al., 2003; Chen et al., 2006). But unlike *N. gonorrhoeae*, the proteins making up the peudopilus are not usually expressed in the cell (Van Sinderen et al., 1995).

Instead, they are present on the cell membrane only when competence is turned on. So competence in this species is a tightly regulated process.

Two quorum sensing systems, mediated by two peptides, ComX and CSF (competence-sporulation factor), are found to be responsible for the competence development in *B. subtilis* (Figure 3). A ComX precursor is synthesized as an inactive 55-residue peptide in the cell and processed to be a 10-a.a. active peptide of ComX when it is exported by a membrane transporter, ComQ. It is believed that the cleavage happens at the C-terminal of pre-ComX and a tryptophan residue is isoprenylated (Ansaldi et al., 2002). The post-translationally modified ComX can activate a typical two-component transduction system (TCTS), the sensor ComP and response regulator ComA, to up-regulate the expression of ComS. ComP is a histidine kinase. It can be auto-phosphorylated upon binding with ComX and transfer a phosphate group to a conserved aspartate residue of ComA. The phosphorylated ComA acts as a transcription factor, binding to *comS* promotor and enhancing its expression. On the other hand, the intracellularly synthesized pre-CSF is exported and processed to be a penta-peptide by a cleavage of the five residues at the C-terminal of pre-CSF (Solomon et al., 1996). Unlike ComX which is sensed extracellularly by a TCTS, CSF is taken up into the cell by the oligopeptide permease, Opp. The imported CSF is thought to inhibit a ComA-specific phosphatase, RapC, thereby facilitating the phosphorylation of ComA and promoting the production of ComS (Solomon et al., 1996; Lazazzera et al., 1997). An interesting observation is that only low levels of CSF are able to enhance *comS* expression; high levels of CSF inhibit *comS* expression instead (Solomon et al., 1996).

ComS is a 46-a.a. protein that is able to bind to the N-terminus of MecA. MecA is the adaptor that directs ComK to the protease machine ClpCP, with its N-terminus interacting with ComK and C-terminal interacting with ClpC (D'Souza et al., 1994; Persuh et al., 1999). ComS binding to the N-terminus of MecA blocks its interaction with ComK and, therefore, increases the accumulation of ComK
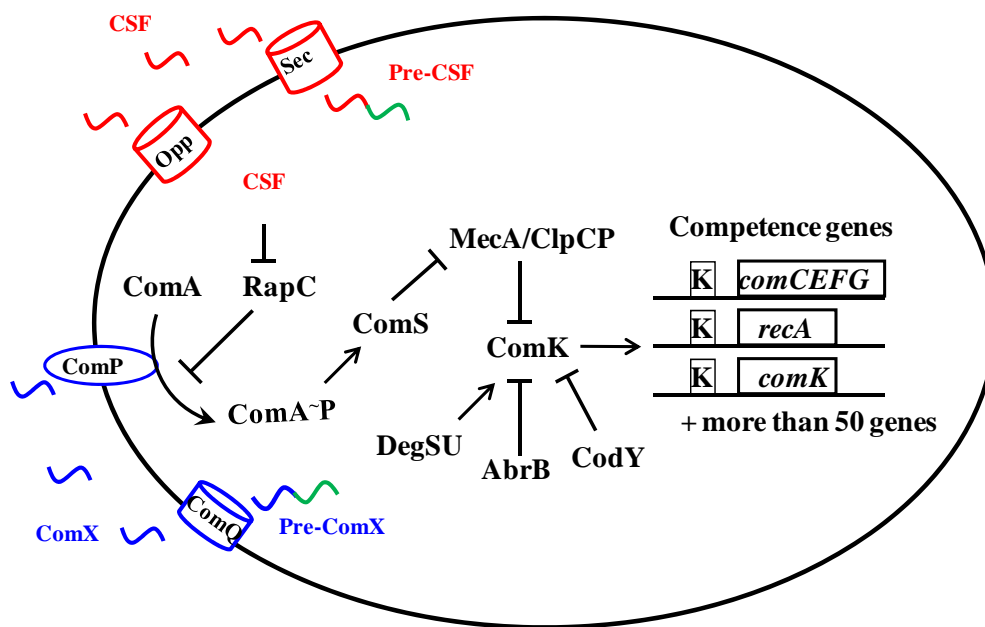
Figure 3 Competence regulation in *Bacillus subtilis*. Competence in B. subtilis is controlled by two related quorum sensing systems mediated two peptides: CSF and ComX. When ComX reaches certain threshold in the environment, it is able to activate a two-component transduction system (ComPA) via auto-phosphorylation. ComP is a histidine kinase and ComA is its response regulator. On the other hand, the matured CSF enters the cell and inhibits a ComA-specific phosphotase, RapC, therefore contributing to ComA phosphorylation. The phosphorylated ComA acts as a transcription factor, binding to *comS* promotor and enhancing its expression. ComS can bind to the N-terminus of MecA and, therefore, protect ComK from the degradation of MecA/ClpCP. Eventually, the increased ComK is able to recognize a conserved sequence, K-box, in front of the promoters of competence genes and up-regulate their expressions, including *comK* itself. The recognition of ComK to K-box is subject to regulations, positively or negatively, by many factors, such as DegSU, AbrB, and CodY.

in the cell. ComK is a competence transcription factor with 192 amino acids. It can bind to a ComK-box (AAAAN$_5$TTT motif) which is upstream of the promoters of competence genes, including *comK* itself, and turn on their expressions like an enhancer (Van Sinderen et al., 1995; Hamoen et al., 1998).

ComK is the central player of the competence development. There are some other auxiliary regulatory proteins also involved in *comK* transcription. The ComK-box is always present in two copies in front of competence genes, with variant distances in the between. The specific sequence between the two ComK-boxes upstream of *comK* makes it to be able to be recognized and regulated by DegSU, AbrB, CodY and others (Hamoen et al., 2000). DegS and DegU together make up a two component regulatory system in *B. subtilis*, in which DegS functions as the sensor kinase and DegU works as the response regulator (Kunst et al., 1988). DegU can recognize and bind to specific sequence between the ComK-boxes of *comK* gene, which enhances the binding of ComK to the ComK-boxes and upregulates *comK* expression. However, AbrB and CodY are transcriptional repressors of *comK*. Binding of AbrB and CodY to the ComK-box region blocks the accessibility of ComK to the promoter region, therefore inhibiting the initiation of *comK* transcription (Hahn and Dubnau, 1991; Hoa et al., 2002).

Taken together, competence development in *B. subtilis* is regulated at multiple levels. Two quorum sensing systems work together to allow the phosphorylation of ComA. Phosphorylated ComA promotes the transcription of ComS. ComS inhibits the proteolysis of ComK by MecA/ClpCP and increases its stability at protein level. DegU alleviates the repression of *comK* expression imposed by AbrB and CodY and upregulates *comK* at messenger level. Finally, when sufficient amount of ComK is accumulated in the cell, the expression of competence genes is turned on.

1.4.4 Competence regulation in *Streptococcus mutans*

Streptococcus mutans is also Gram-positive bacteria which is commonly found in the human oral cavity and is a main contributor to tooth decay. Competence development has been found to be related

to three sets of gene expression in two quorum-sensing systems: CSP-induced upstream regulators, XIP-induced proximal regulators and ComX-induced late genes (Figure 4).

The CSP (competence stimulating peptide) precursor, ComC, is synthesized in the cytoplasm and processed to be a double-glycine signal peptide, CSP, when it is exported though the membrane transporter NlmTE (van der Ploeg, 2005; Kreth et al., 2007). NlmTE is homolous to BlpAB in *S. pneumoniae*, which is involved in production of bacteriocins. A mechanism of quorum sensing is employed to transmit the signal of CSP into the cell by a TCTS: ComDE. ComDE is orthologous to BlpRH in *S. pneumoniae* with ComD serving as the auto-phosporylatable histidine kinase and ComE as the response regulator. Like BlpRH in *S. pneumoniae*, ComDE functions in the regulation of bacteriocin production, not in the regulation of competence (A. de Saizieu et al., 2000; Martin B., 2006). Phosphorylated ComE can regnized a specific sequence in front of a bunch of genes and enhance their expression. One of them is cipB, which codes for mutacin V in *S. mutans*. It is believed that CipB, along with other bacteriocins and memberance regulators, establishes an undefined the link to *comR* expression (Lemme et al., 2011).

Another quorum sensing system is mediated by the secretion of the signal peptide, XIP (ComX-inducing peptide). XIP is processed from its precursor ComS with the cleavage at C-terminal. The matured XIP contains seven amino acid residues, with a double-tryptophan motif at the C-terminal for signaling (Mashburn-Warren et al., 2010). What's different with CSP is XIP is recruited into the cytoplasm to transmit the signal by the oligopeptide permease transporter, Opp. The target that XIP binds is ComR, which is a unique regulator of the Rgg family (Sanders et al., 1998; McIver 2009). Rgg gene was oringally found as regulator gene of glucosyltransferase that works by recognizing a specific sequence in the promoter region and enhancing gene expression (Sulavik et al., 1992). The binding of
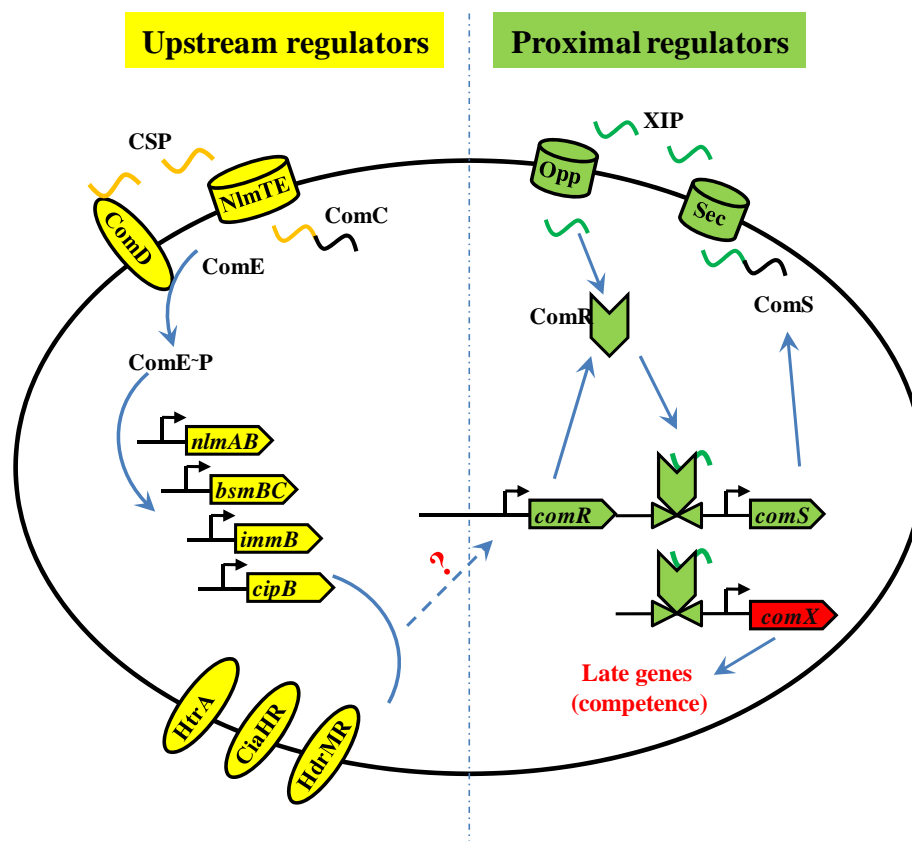
Figure 4 Competence regulation in *Streptococcus mutans*. Competence in *S. mutans* is believed to be controlled by upstream and proximal regulators. CSP, which is processed from endogenous ComC, mediates the upstream regulators. A mechanism of quorum sensing is employed to transmit the signal of CSP into the cell by a two-component system ComDE, where ComD serves as the auto-phosporylatable histidine kinase and ComE as the response regulator. Phosphorylated ComE can recognized a specific sequence in front of several genes and enhance their expression. One of them is *cipB*, which codes for mutacin V in *S. mutans*. It is believed that CipB, along with other bacteriocins   and memberance regulators, establishes an undefined the link to *comR* expression. ComR can be activated by XIP. XIP is processed from its precursor ComS and recruited into the cytoplasm via Opp. The activated ComR-XIP can recognize a specific sequence in the promoter region of several genes and enhance their expression, including XIP precursor *comS* and *comX*. ComX is an alternative sigma factor that is able to turn on expressions of more than 20 late genes that are involved in competence development.

XIP activates ComR and makes it able to recognize a reverse-repeat sequence in front of *comS*, therefore forming a positive regulatory loop to increase the production of ComS and XIP. This reverse-repeat sequence was also found in front of *comX* gene. ComX is an alternative sigma factor that is able to turn on expressions of more than 20 late genes that are involved in competence development.

1.4.5 Competence regulation in *Streptococcus pneumoniae*

One of the most remarkable features of *S. pneumoniae* is its ability to be naturally transformed. Competence in this species has been found to be a transient state in which competence suddenly occurs and rapidly disappears lasting around 30 minutes during the mid-log phase of growth. It is tightly regulated and controlled. Natural genetic transformation in *S. pneumoniae* entails cells becoming competent which is under the control of a quorum sensing system, actually a two-component signal transduction system (TCSTS). Competence development depends on many factors, such as cell density, pH, and the nutrient conditions. (Chen and Morrison, 1987; Tomasz and Hotchkiss, 1964). Generally, competence development starts with the synthesis of a polypeptide ComC, the precursor of competence stimulating pheromone (CSP) (Håvarstein et al., 1996). In *S. pneumoniae*, comC is expressed at a basal level to produce a 41-a.a. long peptide, which belongs to a class of proteins known as bacteriocin characterized by having a double-glycine-type leader peptide. The double-glycine region of ComC services as a cleavage site producing the 17-a.a. long peptide, CSP (EMRLSKFFRDFILQRKK) (Havarstein et al., 1995). This process happens when ComC is transported out through a cell membrane transporter ComA and its accessory partner ComB. It is believed that ComA, an ATP-binding cassette transporter and essential for competence development, makes the cleavage (Havarstein et al., 1995.). When CSP is accumulated to a certain level in the environment outside the cell, it is sensed by the extracellular domain of ComD, a histidine kinase. ComD is then auto-phosphorylated and its
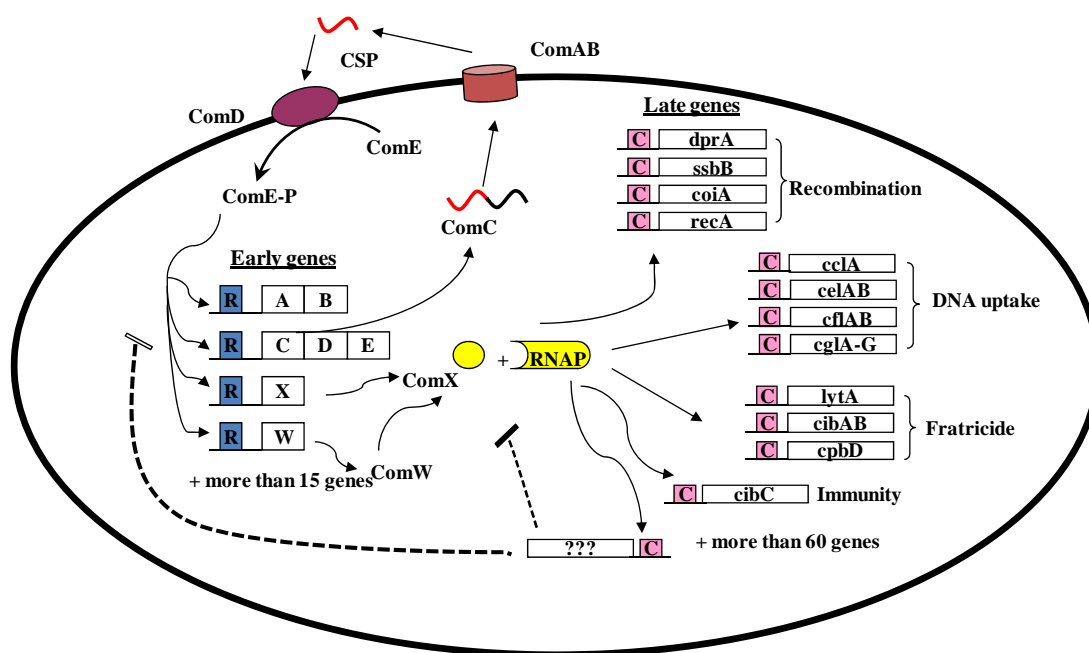
Figure 5 Competence regulation in *Streptococcus pneumoniae*. Competence development in *S. pneumoniae* entails the expressions of two sets of genes, the early gene and late genes, induced by CSP. CSP, which is secreted and processed from an endogenous peptide ComC, can activate the two-component system: histidine kinase ComD and response regulator ComE. Phosphorylated ComE, recognizing an imperfect direct repeat in front of the promoter of several operons, up-regulates the expression of the early genes, including *comAB*, *comCDE*, comX, and ComW. One early gene is *comX*, encoding an alternative sigma factor that recognizes a promoter sequence, termed "combox". When recruited into core RNAP, the RNAP-ComX complex induces a second round of gene expression, termed the late genes. The late genes involve DNA uptake and processing, recombination, fratricide, immunity and other competence functions, creating competent cells. Another early gene product, ComW, is also required for late gene expression, even though its role remains unclear. It has been speculated that a late gene product might be able to shut off competence.

cytoplasmic domain further phosphorylates its response regulater ComE on the aspartate residue. The activated ComE dissociates from ComD and acts as a transcriptional enhancer up-regulating several genes (Havarstein et al., 1995; Pestova et al., 1996). Therefore, it is through the TCSTS composed of ComD and ComE that the signal is transmitted from outside to inside. The phosphorylated ComE can recognize an imperfect direct repeat (CAnTT-16-CAnTT) which is found in front of several genes and enhance their expressions. These gene are termed the early genes (Ween et al., 1999), including *comAB*, *comCDE*, *comX*, *comW*, etc. The expressions of *comAB* and *comCDE* actually form a autocatalytic loop and result in the secretion of more and more CSP into the extracellular environment. The up-regulation of *comX* and *comW* results in the accumulation of ComX and ComW in cytoplasm and turns on another set of gene expression, which is termed the late genes (Alloing et al., 1998; Peterson et al., 2000; Peterson et al., 2004). ComX is an alternative sigma factor and can be recruited into RNA polymerase to recognize a non-canonical promoter sequence (TACGAACA), named combox or cinbox, which are found in front of late genes (Luo, 2003). ComW has been found to be required for competence development and proved to have dual functions in protecting and activating *comX* (Sung and Morrison, 2005.). The late genes under the control of ComX involve in DNA uptake and processing, recombination, fratricide, immunity and others (Stephen and Bassler). One of the late gene products, DprA, has shown to be able to shut off competence probably by interacting with ComE and inactivating its function as enhancer. This will be further discussed in Chapter 4.

1.5 Significances of competence studies

Currently, two major significances have been proposed as to why bacteria spend so much energy to develop and maintain the state of competence. One is for uptake of nutrition and another is for exchange of genetic materials. It has been suggested that the uptake DNA strands could be a good source of energy and building blocks for the synthesis of new DNA (Finkel and Kolter 2001; Redfield,

2001). More importantly, competence development enables bacteria to exchange genetic material with the kin of this species and, therefore, increase its ability to adapt to the changing environments. It is believed that the major reason why *S. pneumoniae* is becoming drug-resistant and able to evade the human immune system is due to its ability to become competent for natural transformation (Hakenbeck, 1999). Also, as a type of stress responses, competence can cross-talk with many other stressful processes, such as competence and nutrient depletion, competence and biofilm formation (Trappetti et al., 2011), competence and production of bacteriocin (Kreth et al., 2005), competence and antibiotic treatment (Prudhomme et al., 2006), etc. All these aspects about competence make it an important biological process to study, either for the purpose of deepening our understanding of the importance of horizontal gene transfer during evolution, or for the purpose of improving our knowledge of microbial pathogenesis and finding out new strategies against infectious diseases.

1.6 Project objectives

When I joined the lab, the process of competence buildup had been well established, but the mechanism of competence shutoff remained obscure. So the major objective of my studies was to figure out how competence is terminated. ComX and ComW are the central player of competence development *S. pneumoniae*. Their degradations were first suspected to be the reason for competence termination. But later evidence implied this may not be the case because competence still disappeared rapidly, even though the level of ComX remained high in the cell. (Luo et al., 2003). Another interesting study showed that competence still shut off when early gene expression was bypassed by ectopically expressing *comX* and *comW*. These data pointed toward the possibility of ComX-dependent late gene product(s) that might turn off competence. Basically, there are two groups of late genes induced by ComX (Figure 6): transformation-essential genes and transformation-nonessential genes.
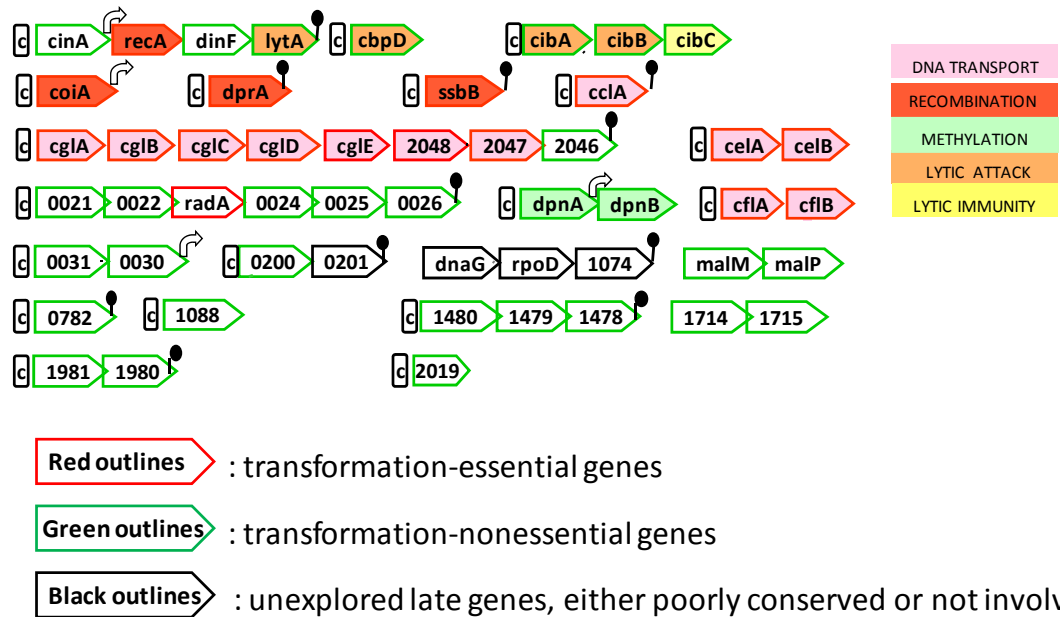
Figure 6 Organization of gene clusters in late regulons of the CSP response in *S. pneumoniae*. Genes described by numbers are unknown genes**.** C in box represents combox. Filled colors suggest the functions of proteins, as indicated to the right. Red border colors indicate essential in transformation. Green border colors indicate not essential in transformation. Black borders indicate not explored. Bent open arrows represent constitutive promoters. Lollipops suggest transcription terminators.

Considering there are still many transformation-nonessential genes without any known functions, it is possible these unknown late genes may be involved in competence regulation. Maybe there are some more branched processes associated with competence development we just do not understand. Therefore, studying possible roles of these unknown late genes in competence development was on the top of the list of my goals. Since previous lab work had shown that any single mutation of them almost had no effect on competence, we wanted to investigate if there is any redundancy existed or if they are just dispensable during competence development. To achieve this goal, we need to have an efficient and reliable tool to make multiple gene deletions without being subject to limited gene markers (Em^R, Cm^R, Kan^R, Str^R, Spc^R) used in *S. pneumoniae.*

My first aim in this Ph.D program was to set up the markerless multiple gene deletion system in the application of *Streptococcus pneumonia.* Cre-loxP recombination system has been used in many species to make markerless gene deletion. Incorporating it into *S. pneumoniae* should be valuable. Some new design was incorporated to make its application more convenient and efficient. Once the system is established, I would try to make multiple deletions of the unknown but highly conserved late genes to investigate their functions. By doing that, I would be able to answer if they are redundant during transformation and if any combination of them has a role in competence termination. This part will be discussed in Chapter 3.

Another group of late genes is the transformation-essential genes. Even their functions had been well documented, their possible roles in competence termination had not been explored simply because they are required for competence buildup. Any disruption of them would result in no transformation, therefore there is no way to look at the competence phenotype directly. But I suspected that some of them might be able to turn off competence. Because during the period of competence, many bacteriocins are produced to kill non-competent pneumococcal cells and other symbiotic species. It needs to be ended

in time to make sure the whole population could exit this state and go back to normal life. I believe this vital role is more likely to be performed by a transformation essential gene rather than a transformation-nonessential gene. And there is an indirect way to study competence by putting an reporter gene in a late gene operon. Instead of looking at competence kinetics directly, the late gene expression patterns can be examined to indicate competence development, since late gene expression has been proved to be a reliable indicator of competence development. To unravel the mystery of competence regulation, I decided to explore these transformation-essential genes to see if any of them has a role in turning off competence. Once the late gene repressor is identified, I would move forward to characterize how it shuts off competence and what target(s) it might work on. This part will be elaborated in Chapter 4.

ComX and ComW are the central players of competence development. But how they work in concert to turn on competence has not been clarified. Even though it had been suspected that they might interact each other, endeavors of obtaining evidence to verify their interaction were not successful in the lab. So another goal of my project was to investigate to see if these two protein are able to interact or not. This part will be detailed in Chapter 5.

Chapter II. Materials and Methods

2.1 Bacterial strains, plasmids and culture methods

The bacterial strains and plasmids used in this study are listed in Table I and II below. Detailed construction strategies are described in Appendices. All strains were stored frozen at -80ºC. For *S. pneumoniae* strains, cells were grown to OD_{550nm} at 0.1 in CAT complete medium, and stored with 15% sterile glycerol in final concentration. When used, the frozen stocks were inoculated to fresh medium with a 100-fold dilution. For *E. coli* strains, cells were grown to late exponential phase in LB, and stored with 20% sterile glycerol in the final concentration. When used, the frozen stocks were streaked onto LB agar plates with appropriate antibiotics, and a single colony was then inoculated into fresh medium. For yeast strains, cells were grown in YPD or synthetic selective media, and stored with sterile glycerol at 20% final concentration. When used, frozen stocks were streaked onto YPD agar plate, and a single colony was then inoculated into fresh YPD or synthetic selective media.

2.2 Media and drugs

CAT medium (Casein hydrolysate yeast extract) was used for growing *S. pneumoniae* cells (Morrison *et al.*, 1983). CAT base contained 0.1% Bacto Yeast Extract, 0.5% Bacto Tryptone, 0.5% NaCl, 1% ICN Casein Hydrolysate. After completely dissolved, CAT base was autoclaved with Liq40 program for sterilization. When used, CAT base was supplemented with 1/30 volume 0.5 M K_2HPO_4 and 1/100 volume 20% glucose to form CAT broth. CAT broth with 1.5% Bacto agar was used for plating.

LB broth medium (Luria-Bertani), was used for growing *E. coli* cells. LB medium contained 1% Bacto Tryptone, 0.5% Bacto Yeast Extract and 0.5% NaCl (Sambrook *et al.*, 2001). LB medium plus 1.5% Bacto agar was used to make LB agar plates.

TABLE I. BACTERIAL STRAINS USED IN THIS STUDY

Strain	Description	Source (a)
S. pneumoniae		
CPM1	CP1250, but *rpoC*::C-His$_{10}$::pEVP3; Cm^R	Lee, et al. 1999
CPM2	CP1250, but Δ*comX1*::PcEm; Em^R	Lee, et al. 1999
CPM7	CP1250, but *ssbB$^-$*::*lacZ*::*ssbB$^+$*; SsbB$^+$ Sm^R Cm^R	Lee, et al. 1999
CP1250	Rx, but *cps3D hex cps3D malM511 str-1 bgl-1*; Hex$^-$ Mal$^-$ Sm^R Bga$^-$	Pestova, et al.1996
CP1275	CP1250, but Δ*cbpD*::PcKan; Kan^R	Peterson, et al.2004
CP1279	CP1250, but Δ*cibABC*::PcKan; Kan^R	Peterson et al. 2004
CP1288	CPM2, but *comX2*::His6::pEVP3; Cm^R Em^R	Luo, et al.2003
CP1333	CP1250, but Δ*cglEFG*::PcKan; Kan^R	Sung, et al.2005
CP1334	CP1250, but Δ(spr1630spr1631)::P$_C$Kan; Sm^R Kan^R	Peterson, et al.2004
CP1344	CP1250, but Δ*clpC*::P$_C$Tet; Tet^R	Piotrowski, et al.2009
CP1359	CP1250, but Δ*clpP*::PcTet; Tet^R	Piotrowski, et al.2009
CP1389	CP1250, but Δ*dprA*::PcKan; Kan^R	Desai, et al.2006
CP1415	CP1250, but Δ*comA*::PcErm; Em^R	Morrison, et al.1984
CP1500	Rx, *hex nov-r1, byr-r, ery-r1,ery-r2, str-1*;Nov^R Ery^R Sm^R,	Cato, et al.1968
CP1793	CP1250, but Δ*coiA*::PcKan; Kan^R	Desai, et al.2006
CP1851	CP1250, but Δ*clpE*::PcErm; Em^R	Piotrowski, et al. 2009
CP1862	CP1250, but Δ*celAB*::PcKan; Kan^R	Appendix C
CP1863	CP1250, but Δ*cclA*::PcKan; Kan^R	Appendix C
CP1868	CP1250, but Δ*cglABCD*::PcKan; Kan^R	Appendix C
CP1869	CP1250, but Δ*cflAB*::PcKan; Kan^R	Appendix C
CP1890	CP1250, but *ssbB$^-$*::*lacZ*::*ssbB$^+$*,Δ*clpP*::PcTet; Cm^R Tet^R	CPM7 x CP1359
CP1894	CP1250, but *ssbB$^-$*::*lacZ*::*ssbB$^+$*,Δ*dprA*::PcKan; Cm^R Kan^R	CPM7 x CP1389
CP1895	CP1250, but *ssbB$^-$*::*lacZ*::*ssbB$^+$*,Δ*clpP*, Δ*dprA;* Cm^R Kan^R Tet^R	CP1890 x CP1389
CP1961	CP2000, but *aga*::*comX*::*comW*, *ssbB$^-$*::*lacZ*::*ssbB$^+$*; Cm^R Kan^R	CP1896xCP2000xCPM7
CP1962	CP1961, but Δ*clpE*::PcErm; Em^R	CP1961 x CP1851
CP1963	CP1961, but Δ*clpC*::PcTet; Tet^R	CP1961 x CP1344
CP1902	CP1961, but Δ*clpC*::P$_C$Tet*,* Δ*clpE*::PcErm; Tet^R Em^R	CP1962 x CP1344
CP1932	CP1902, but Δ*dprA*::PcKan; Kan^R	CP1902 x CP1389
CP1939	CPM7, but Δ*ssbB*::erm-P$_{fcsK}$::(pEVP3):: *ssbB$^+$*; Sm^R Em^R Cm^R	Appendix C
CP2000	CP1250, but Δ*cps*; Hex$^-$ Mal$^-$ Cps$^-$ Sm^R Bga$^-$	Appendix C
CP2009	CPM1, but Δ*clpE*::PcEm; CmR Em^R	CPM1 x CP1851
CP2026	CP2009, but Δ*clpC*::PcTet; Tc^R Cm^R Em^R	CP2009 x CP1344
CP2027	CP2026, but Δ*comW*::PcKan; Tc^R CM^R Em^R Kan^R	CP2026 x CP1376
CP2052	CP2000, but Δ*ssbB*::erm-P$_{fcsK}$::(pEVP3):: *ssbB$^+$*; Cps$^-$ Sm^R Em^R Cm^R	CP2000 x CP1939
CP2053	CP2000, but Δ(spr1630spr1631)::PcKan::lox72, Str^R Kan^R	CP2000 x CP2056
CP2054	CP2000, but Δ(spr1630spr1631)::PcKan::Cheshire, Str^R Em^R	CP2000 x CP2055
CP2055	CP1334, but P$_C$Kan::Cheshire; Sm^R Em^R	Appendix C
CP2056	CP1334, but PcKan::lox72, Sm^R Kan^R	Appendix C

CP2057	CP1288, but $\Delta clpP$::PcTet; Cm^R Em^R Tet^R	CP1288 x CP1359
CP2062	CP2000, but Δ(spr1630spr1631)::P_CKan::Cheshire; Sm^R Em^R	CP2000 x CP2055
CP2108	CP2000, but $ssbB^-$::lacZ::$ssbB^+$, $\Delta comA$::PcErm; Sm^R Cm^R Em^R	CP2000xCPM7xCP1415
CP2109	CP2108, but $\Delta cbpD$::PcKan; Kan^R	CP2108 x CP1275
CP2110	CP2108, but $\Delta cibABC$::PcKan; Kan^R	CP2108 x CP1279
CP2111	CP2108, but $\Delta coiA$::PcKan; Kan^R	CP2108 x CP1793
CP2112	CP2108, but $\Delta cglEFG$::PcKan; Kan^R	CP2108 x CP1333
CP2113	CP2108, but $\Delta dprA$::PcKan; Kan^R	CP2108 x CP1389
CP2114	CP2108, but $\Delta cclA$::PcKan; Kan^R	CP2108 x CP1863
CP2115	CP2108, but $\Delta cflAB$::PcKan; Kan^R	CP2108 x CP1869
CP2116	CP2108, but $\Delta celAB$::PcKan; Kan^R	CP2108 x CP1862
CP2117	CP2108, but $\Delta cglABCD$::PcKan; Kan^R	CP2108 x CP1868
CP2118	CP1250, but $\Delta radA$::PcSpc; Spc^R	D. Ringus
CP2119	CP2108, but $\Delta radA$::PcSpc; Spc^R	CP2108 x CP2118
CP2125	CP2108, but $\Delta clpP$::PcTet; Tet^R	CP2108 x CP1359
CP2126	CP2125, but $\Delta cbpD$::PcKan; Kan^R	CP2125 x CP1275
CP2127	CP2125, but $\Delta cibABC$::PcKan; Kan^R	CP2125 x CP1279
CP2128	CP2125, but $\Delta cglEFG$ ::PcKan; Kan^R	CP2125 x CP1333
CP2129	CP2125, but $\Delta dprA$::PcKan; Kan^R	CP2125 x CP1389
CP2130	CP2125, but $\Delta coiA$::PcKan; Kan^R	CP2125 x CP1793
CP2131	CP2125, but $\Delta celAB$::PcKan; Kan^R	CP2125 x CP1862
CP2132	CP2125, but $\Delta cclA$::PcKan; Kan^R	CP2125 x CP1863
CP2133	CP2125, but $\Delta cglABCD$::PcKan; Kan^R	CP2125 x CP1868
CP2134	CP2125, but $\Delta cflAB$::PcKan; Kan^R	CP2125 x CP1869
CP2135	CP2125, but $\Delta radA$::PcSpc; Spc^R	CP2125 x CP2119
CP2139	CP2108, but $\Delta ssbB$::PcKan; Kan^R	Appendix C
CP2140	CP2125, but $\Delta ssbB$::PcKan; Kan^R	Appendix C
CP2143	CP2108, but ΔPcom-$cinA$::PcKan; Kan^R	Appendix C
CP2144	CP2125, but ΔPcom-$cinA$::PcKan; Kan^R	Appendix C
CP2154	CP2057, but ΔpEVP3:: PcSpc; Spc^R Em^R Tet^R	Appendix C
CP2155	CP2057, but ΔcomW:: PcKan; Kan^R Em^R Tet^R	CP2057 x CP1376

E. coli

BL21	F-ompT $hsdS_B$ (r_B-m_B-) gal dcm (DE3)	Invitrogen
DH5α	F-recA1, endA1 hsdR17 phoA supE44 thi-1 gyrA96	Invitrogen

Yeast

NSY468	MATa, trp1-901, leu2-3, l 12, ura3-52, his3-200, gal4Δ, gal80Δ GAL2-ADE2, LYS2::GAL1-HIS3, met2::GAL7-lacZ	James, et al.1996
NSY752	MATα, trp1-901, leu2-3, l 12, ura3-52, his3-200, gal4Δ, gal80Δ GAL2-ADE2, LYS2::GAL1-HIS3, met2::GAL7-lacZ	James, et al.1996

(a) Crosses are indicated as recipient x donor genomic DNA

TABLE II. PLASMIDS USED IN THIS STUDY

Plamid	Description	Source
pEVP3	lacZ reporter insertion vector	Pestova, et al.1996
pMAL-c2x	protein expression vector with MBP	NEB, Inc.
pAP01	pMAL-c2x derivative, carrying *comW*	Piotrowski, et al.2010
pXPL01	pCR-T7/CT-TOPO derivative, carrying *comX*	Luo, et al. 2003
pACT2	shuttle vector for yeast 2-hybrid, carrying Gal4 AD	Clontech Labs, Inc.
pACT2-*comD*	pACT2 derivative, carrying AD-*comD* fusion	Appendix D
pACT2-*comE*	pACT2 derivative, carrying AD-*comE* fusion	Appendix D
pACT2-*comX*	pACT2 derivative, carrying AD-*comX* fusion	Appendix D
pACT2-*comW*	pACT2 derivative, carrying AD-*comW* fusion	Appendix D
pACT2-*dprA*	pACT2 derivative, carrying AD-*dprA* fusion	Appendix D
pACT2-*recA*	pACT2 derivative, carrying AD-*recA* fusion	Appendix D
pGBDUC2	shuttle vector for yeast 2-hybrid, carrying Gal4 BD	James, et al.1996
pGBDUC2-*comD*	pGBDUC2 derivative, carrying BD-*comD* fusion	Appendix D
pGBDUC2-*comE*	pGBDUC2 derivative, carrying BD-*comE* fusion	Appendix D
pGBDUC2-*comX*	pGBDUC2 derivative, carrying BD-*comX* fusion	Appendix D
pGBDUC2-*comW*	pGBDUC2 derivative, carrying BD-*comW* fusion	Appendix D
pGBDUC2-*dprA*	pGBDUC2 derivative, carrying BD-*dprA* fusion	Appendix D
pGBDUC2-*recA*	pGBDUC2 derivative, carrying BD-*recA* fusion	Appendix D

YPD was the complete medium used for growing all kinds of yeast strains, both haploids and diploids. YPD contained 1% Bacto Yeast Extract, 2% Bacto Peptone, and 2% Dextrose. YPD broth plus 1.5% Bacto agar was used to make YPD plates.

SD selective medium (Synthetic Defined) was used for growing selected yeast strain. 1 L SD medium contained 7.1 g YNB super powder (yeast nitrogen base and most amino acids), 20 g glucose, and 40 ml 25X appropriate specific amino acid dropout solutions. 20 g of agar was added to SD broth medium to pour SD agar plates. 400 ml dropout solutions were made by selectively mixing some of the eight amino acids (0.3 g tryptophan, 2.0 g threonine, 0.2 g methionine, 0.3 g lysine, 0.4 g adenine, 0.8 g leucine, 0.2 g uracil, and 0.2 g histidine), filter-sterized and stored at 4ºC. For example, the dropout-leu-ura contains all the six amino acids but not leucine and uracil, and so forth.

Concentrations of antibiotics used for selection of *S. pneumoniae* were: erythromycin (Em), 0.2 µg/ml; chloramphenicol (Cm), 2.5 µg/ml; kanamycin (Kan), 200 µg/ml; tetracycline (Tet), 0.25 µg/ml; spectinomycin (Spc), 100 µg/ml; streptomycin (Str), 150 µg/ml. Ampicillin used for selection of *E. coli* was 100 ng/ml.

2.3 Culture growing conditions

For growing *S. pneumoniae*, cultures were started by inoculating complete CAT medium plus 10 mM HCl with 1/100 volume of a frozen stock of cells (OD_{550}=0.1) and incubated at 37ºC in 10 ml with 18 mm by 150 mm, or 4 ml with 13 mm by 100 mm tubes for monitoring optical density. OD measurements were taken every 30 minutes to monitor the growth. When the cell density reached OD_{550}=0.02~0.04, it was induced to competence with $CaCl_2$ (0.5 mM), BSA (0.002%), and CSP (250 ng/ml). Samples were taken periodically from the culture for the purposes of various analysis.

For growing *E. coli*, cultures were started by inoculating LB medium with 1/100 volume of a frozen stock of cells and incubated at 37ºC with the shaking of 220 rpm. OD measurements were taken every 30 minutes to monitor the growth.

For growing yeast, cultures were started by inoculating YPD or SD selective medium with 1/100 volume of a frozen stock of cells and incubated at 26ºC with the shaking of 220 rpm. Cultures usually were grown overnight for subsequent usages.

2.4 DNA extraction

2.4.1 Extraction of genomic DNA from *S. pneumoniae*

Genomic DNA of Pneumococcus was purified by chloroform extraction and ethanol precipitation based on the method described previously (Marmur, 1961). 300 ml CAT plus phosphate plus antibiotic was inoculated with a 1/300 volume of a frozen stock of cells (OD_{550}= 0.1) and grown overnight at 37°C. Glucose was added to 0.2% and the culture was grown to OD550=0.6. The culture was chilled on ice and EDTA was added to 5 mM. After centrifugation with 9,000xg for 20 minutes at 4°C, the cell pellet was resuspended in 1/10 volume of Buffer A (0.05 M Tris-HCl pH7.6, 10 mM EDTA, 0.1 M NaCl). The resuspended pellet was centrifuged with 9,000xg for minutes at 4°C. The pellet was resuspended in 10ml Buffer A plus 0.5% Triton X-100 and incubated at 37°C for 10 minutes. After that, SDS was added to 0.2% and incubated at 65°C for 20 minutes. The primary cell lysate was frozen at -20°C overnight. The lysate was allowed to thaw and an equal volume of chloroform plus 4% iso-amyl alcohol was added and the mixture was shaken vigorously. The mixture was centrifuged with 9,000xg for 30 minutes at 4°C. The supernatant was taken out carefully and added to an equal volume of chloroform plus 4% iso-amylalcohol and repeat above step. The supernatant with genomic DNA was taken out. The addition of RNase A to the supernatant was optional. 1/10 volume of 3M sodium acetate was then added to the supernatant and the solution was inverted several times to mix well. The

supernatant was added with 2.5 volumes ice cold ethanol. The precipitated DNA was spun onto a glass rod and allowed to dry in the air. The dried DNA was re-dissolved in 5 ml TE Buffer (10 mM Tris-HCl pH7.6, 1 mM EDTA).

2.4.2 Extraction of plasmid DNA from *E. coli*

Plasmid DNA in E. coli was purified with QIAprep Miniprep plasmid DNA purification kit (Qiagen Inc.). Basically, a single colony was picked from a freshly grown selective streak plate and inoculated into 5~10 ml LB medium containing the appropriate selective antibiotic. The culture was incubated for 12~16 h at 37°C with vigorous shaking. The cells were harvested by centrifuging at 8,000 rpm at 4°C for 3 minutes. The pellet was resuspended with 250 µl resuspending buffer (P1) and transferred to a micro-centrifuge tube. 250 µl lysis buffer (P2) was then added to lyse the cells by thoroughly mixing the solution. 350 µl neutralizing buffer (N3) was applied to precipitate proteins and lipids followed by centrifuging at 13,000 rpm at RT for 10 minutes. The supernatant containing plasmid DNA was applied onto a QIAprep spin column and centrifuged at 10,000 rpm at RT for 30 seconds. The flow-through was discarded and the column was washed 750 µl wash buffer (PE) by centrifugation. The flow-through was discarded and the column was spun one more minute to get rid of wash buffer residue. Then the column with absorbed plasmid DNA was placed onto a clean micro-centrifuge tube and 30~50 µl elution buffer (EB) was added onto the filter of the column. The purified DNA was eluted out by centrifugation at 10,000 rpm at RT for 1 minute after sitting 5 minutes on the bench.

2.5 Molecular cloning

2.5.1 PCR amplification

Primers used in this study are listed in Table III. They were designed by using Clone Manager 9.0 and synthesized by Operon Inc (Alameda, CA). The criteria of primer design followed the default values in the software, where the criteria for dimers were strictly followed and some criteria such as runs

and repeats were more flexible. The oligonucleotide primers were dissolved in TE buffer (10 mM Tris-HCl pH7.6, 1 mM EDTA) at a final concentration 100 pmol/μl, and stored at -20°C. The working stocks were made by diluting the original stocks two-fold with water. Templates used in PCR were purified DNA, or even the stock cell cultures. Unless otherwise specified, 10 ng of template DNA or 1 μl of stock cell culture (OD_{550}=0.1 for *S. pneumoniae*), 0.5 μl of each primer, 25 μl of Fermentas PCR Master Mix (2x, containing recombinant *Taq* DNA polymerase, dNTPs, Mg^{2+}, salt buffer), and 23 μl water were adjusted to 50μl in total volume. PCR cycles (normally 30 to 35 cycles) varied dependent on the annealing temperature of the primers and the size of the PCR products. The annealing temperature was calculated based on the primer Tm values, often 5°C to 10°C below the lower Tm of the two primers (Sambrook et al., 2001). The extension time varied with the size of the applicon, usually 1 minute per kb. After PCR, 10 μl of the reaction mixture was loaded onto an agarose gel for verifying the PCR product with the correct size. If the PCR product had further use for digestion and ligation, it was treated with the QIAquick PCR Purification Kit (Qiagen Inc.) according to the manufacturer's instructions, and its concentration was determined by using Nanodrop 1000 (Thermo Scientific), or by comparison to a DNA standard ladder within the agarose gel with the Alpha Imager (Alpha Innotech), stained with ethidium bromide.

2.5.2 Endonuclease digestion

DNA fragments used for digestions were usually prepared from PCR amplification or plasmid purification. If it is from PCR amplification, the restriction sites were designed at the 5'-end of the primer, with 2~4 extra nucleotides beyond the restriction site to ensure efficient endonuclease recognition and cleavage. Each restriction enzyme was used following the instructions given by the company. In a digestion reaction with total volume of 50 μl, usually 1-10 μg of DNA plus 1~2 μl restriction enzyme were added and incubated at 37°C for 2 hours. After digestion, the digested DNA

TABLE III. PRIMERS USED IN THIS STUDY

Primer	Location	Sequence, 5' to 3' (a)
AP19	*comW*	ATGTTACAAAAAATTTATGAGCAGATGGC
AP20	*comW*	GATTCTAGATGTCTGAGCCATCAGGCTCTTCC
BVD26	*coiA*	AAACGGGAGTCTATCAAACGTCGTGAGCAA
BVD27	*coiA*	atggatccTGAATTCCCTCCTTTTCTATATCAT
BVD28	*coiA*	atgggcccGAATAGAAAGGATGGAGGAATCTAA
BVD29	*coiA*	GTAGACATCGTACATCTTGAGATCTGAAAT
BVD104	*ssbB*	atggatccTGCCATTTTAAGAATTAAAAAGTC
BVD105	*ssbB*	GACTCTTCGATGGTGATGACACCGTCTTTG
DAM303	*kan*	aagggcccGTTTGATTTTTAATG
DAM304	*kan*	aggatccATCGATACAAATTCCTC
DAM305	*cibABC*	CAAGGACTGACTAGGTAAACAGC
DAM306	*cibABC*	gctaggatccGAGGGCACTCTTGTCTGG
DAM307	*cibABC*	acgagggcccGATAGCAAAAGCAAATAA
DAM308	*cibABC*	CAAGAGGCCGTGTTCTTCGAG
DAM313	*cbpD*	AGCTTTCTCGTGGTGTAGAACAAC
DAM314	*cbpD*	acgaggatccGATCCATTTCCTCTGGAATA
DAM315	*cbpD*	agcagggcccAGGTCTCTGGTAAGTGGTAT
DAM316	*cbpD*	CTCTCAAGGTCGCCCAGCTATG
DAM343	*kan*	GAAGTACATCCGCAACTGTCCATA
DAM345	*kan*	CAGGAGACATTCCTTCCGTATCTT
DAM419	*cglEFG*	CTGTAATTGAGCCTCCGTTACCAATATG
DAM420	*cglEFG*	atggatccGAGTCTGGTTGCTATGATTAGTCT
DAM421	*cglEFG*	atgggcccTTAGCTACCCTCAAGACTTCTTC
DAM422	*cglEFG*	TTGTGCAGACCTACTTGACAGCCTATTATG
DAM563	*dprA*	GATAGAGGCGATAAGCATGGCACATAGTAA
DAM564	*dprA*	atgggcccTGCCATCATTTGATTCAAGAAG
DAM565	*dprA*	ggatccATAACGGCTGGATTACGGCAACCT
DAM566	*dprA*	GATTGGGAACTCGCTTGCGTCCTATGACTGA
DAM659	*celAB*	CTAATTCTGGAGCAGGCGGCCATGTG
DAM660	*celAB*	cgcggatccTTTCAACTGCTTATTTATTTGC
DAM661	*celAB*	acgtgggcccGGAAGGATAAATGTTGTAGATTAG
DAM662	*celAB*	TGAGCCAGCATTTGGCCTGACTGAG
DAM663	*cclA*	TGTTGAGTGGCGACGATAAATAAGG
DAM664	*cclA*	cgcggatccTAGTATAATGGAGAAACATAGATAAG
DAM665	*cclA*	acgtgggcccTTGTTTGATAAAGTCCAATTTC
DAM666	*cclA*	AACAAGCCATTTGGCAGTTTGAGTC
DAM679	*cglABCD*	cgcggatccTCCTCACCTATACTATTCGCAAAG
DAM680	*cglABCD*	TGCAGCGTAGCCATTATTGGTTCAG
DAM681	*cglABCD*	ATCCGTACGAACCCTCGTCACTAAG
DAM682	*cglABCD*	acgtgggcccTGATTTTACTGGAAGCAGTAGTC
DAM683	*cflAB*	cgcggatccAATCATGGAATTTAGGACAATTAAAG
DAM684	*cflAB*	TTCAATCATGCTAAGGGCAATACGG
DAM685	*cflAB*	ACGTGGGCCCTCATAAAAACAAAAATGTTTAG
DAM686	*cflAB*	acgtgggcccTCATAAAAACAAAAATGTTTAG
DAM786	*aga*	AAACTGGGTGGAAGTCTAGAAAGTC
DAM790	*comW*	ctagaattcCTCAACAAGAAATAAACCCCC

DAM791	P_cKan:r*afE*	atcgaattcGGATCCGTTTGATTTTTAATGG
DAM793	P_cKan:r*afE*	AACATCGGTATAGCCAGCACCTTCC
DAM794	*comW*	atcggatccAAAAAAGAAAAGGAGTATTTGA
DAM795	*SP2061*	TCCAGCCTCTGCGCCAGACCTATC
DAM796	*kan*	gcgcggtctcagtatgctatacgaacggtaCATTTATTATTTCCTTCCTCTT
DAM797	*kan*	cgcgggtctcaatacattatacgaacggtaTGATCGAAAAATACCGCTGC
DAM798	*cclA*	CGCATGGGTGCAGGGGATTTCCTC
DAM799	*SP2061*	CGTGAGGTGGAGCAGATG
DAM800	*kan*	gctatacgaacggtaAGCCATTTATTATTTCCTTCCT
DAM801	*kan*	gcgcggtctcagtatGCTATACGAACGGTAAGCC
DAM802	*kan*	cattatacgaacggtATCGAAAAATACCGCTGCGTAAAAG
DAM803	*cclA*	TCGATATCCGCATGGGTG
DAM804	*kan*	gcgcggtctcaataCATTATACGAACGGTATCGAA
DAM805	*kan*	gcatggtctcaggtaAGCCATTTATTATTTCCTTCCTCT
DAM806	*kan*	gcatggtctcatATCGAAAAATACCGCTGCGTAAAAG
DAM807	*SP2061*	CCGTCAAGTGATTGCTAGAG
DAM808	*kan*	gcatggtctcaggtaTTTAGCCATTTATTATTTCCTTC
DAM809	*kan*	gcatggtctcataAAAAATACCGCTGCGTAAAAG
DAM810	*cclA*	CGATATCCGCATGGGTG
DAM811	*oxr*	CAGCGAAACACTGGACTG
DAM812	*ssbB*	gcgcggatccAATTCGAGCTCCCATCAAAC
DAM813	*ermAM*	gcgcggatccTTAGCTCCTTGGAAGCTGTC
DAM814	P_*fcsk*	gcgcgaattcTTTTCTTCTCTCTTCGTCCTTGA
DAM815	*lacZ*	gcgcgaattcTGTGGAAGTTACTGACGTAAG
DAM816	*cat*	GTGCAGGAGCTCGTTATC
DAM817	*kan*	gcatgcatggtctcaggtaTTTAGCCATTTATTATTTCCTTCCTCT
DAM818	*kan*	gcatgcatggtctcataAAAAATACCGCTGCGTAAAAG
DAM819	*kan*	cgtaggatccTTATTATTTCCTTCCTCTTTTCTACAG
DAM820	*kan*	cgtagaattcAATACCGCTGCGTAAAGA
DAM821	*SP2061*	CAACCGCAACTTGACTACC
DAM822	*cclA*	GGCAAGTCGTCCTAATCAC
DAM823	*cps3*	CGCGGATCCTTGGAGTTAGAATAGGGCA
DAM827	*cps3*	GCCTCATCACCAGCCTCAGTAAC
DAM836	*SP2061*	CAACAACCGCAACTTGAC
DAM837	*kan*	cgtaggatccAGCCATTTATTATTTCCTTCCTC
DAM838	*kan*	cgtagaattcGCGTAAAAGATACGGAAGGA
DAM839	*cclA*	AATCACCGCTGGTTTGAC
DAM840	*SP1813*	gcgcatcggctattatcggtaAGCCTTGTCCGAAGAAC
DAM841	*kan*	gcgcggtctcagtatgctatacgaacggtaAGCCATTTATTATTTCCTTCCT
DAM842	*ermAM*	cgcgggtctcaatacattatacgaagttatTTAGCTCCTTGGAAGCTGTC
DAM843	P_*fcsk*	gcgcatcgggtctcaAGTCTTTTCTTCTCTCTTCGTCCTTGATT
DAM844	*cre*	gcgcatgcggtctcAGACTATGTCCAATTTACTGACCGTACAC
DAM845	*cre*	gcgcggtctcatgctatacgaagttaTCTAATCGCCATGTTCCAGCAG
DAM846	*kan*	gcgcggtctcaagcatacattatacgaacggtaTCGAAAAATACCGCTGCGTAAA
DAM847	*SP1809*	gcgctaccgataacgcaTAGCTTCTTGTGCTCTCGTCTT
DAM868	*kan*	ggggacgcgtTGGCTTACCGTTCGTATAG
DAM869	*kan*	ggggccatggTCGATACCGTTCGTATAATGT
DAM896	*lox66*	AGCATGGCATATCAGGAACC
DAM897	*cre*	CGCGCCTGAAGATATAGAAG
DAM898	*lox71*	CACCAGCCAGCTATCAACTC
DAM899	*kan*	CGCAGAAGGCAATGTCATAC

DAM900	Cheshire	ggggagatctTGGCTTACCGTTCGTATAG
DAM901	Cheshire	gactgggcccTCGATACCGTTCGTATAATGT
DAM902	*comB*	GGCCATGGGTACAATCAC
DAM903	*comA*	GAAATTTGGGAAACGTCACTATC
DAM922	*SP2019*	CTTGCAGCTCCTTCCAAC
DAM923	*SP2019*	gcatagatctGCAAAAGAAACAGTTCTGTTTTCA
DAM924	*SP2019*	gcatgggcccTTAATTACTGGTAAAAATGGTACAGGAA
DAM925	*SP2019*	TAGGGCACCTCATGACAAC
DAM926	*SP0022*	ATCCAGCAGCCCTAAGC
DAM927	*SP0022*	gcatagatctGCCATCAGCTGCATCTC
DAM928	*SP0022*	gcatgggcccTTTGCCAAGTGATTGAGTGTG
DAM929	*SP0022*	GCACCGTATCCACCATATGC
DAM930	*comA*	TGGTTAAGGTGGCAGAGG
DAM931	*comA*	gcatagatctGCGCTGTCTGAAGCTTG
DAM932	*comB*	gcatgggcccGGGTTAAGCAGTCAATTGGAATC
DAM933	*comB*	GAGCACGACTTCCAAAGG
DAM934	*ssbB*	gcatgggcccTGAAGAAAGCAGACAAGTAAGC
DAM935	*ssbB*	GGCCTATCTGACAATTCCTG
DAM936	*P_{com}-cinA*	gcatgggcccCGCAGGAATTTTCCTACGATTG
DAM937	*P_{com}-cinA*	CAAGGGACAGAAACCTTAGC
DAM938	*P_{com}-cinA*	gtcaggatccGAGTGGCAGGACCAGATAG
DAM939	*P_{com}-cinA*	GGTGCTCTGCCAAGTATTTC
DAM940	*SP1714-15*	GACGGAAGAGTTGGTCAATG
DAM941	*SP1714-15*	CCATACACTGGAACCCAATC
DAM942	*SP1714-15*	AGGGCAGAAAGGACTAGC
DAM943	*SP1714-15*	CATCCTTTCCCTCCTTATCG
DAM944	*P_{com}-cinA*	TTTCAGACGGCTGTAGGAG
DAM945	*P_{com}-cinA*	GACACCACACCATGTTCTTC
DAM946	*P_{com}-cinA*	CGAATGCGTTACCAAGATCC
DAM947	*P_{com}-cinA*	GGGTGTCCTTCTAGGCTATC
DAM948	*cglABCD*	AAGCAGCCAAGTGGAATG
DAM949	*cglABCD*	CCAGATCTTGACCGACTTC
DAM950	*cglABCD*	CTGGGTTCCTGTGATTATGC
DAM951	*cglABCD*	GCAAGACCACCAACATCTC
DAM952	*cflAB*	GTTCATGCGGTAACAGGAG
DAM953	*cflAB*	GTCTGCCCACATAACAAGC
DAM954	*cflAB*	GGCCGTCCTCTTTAATTGTC
DAM955	*cflAB*	GGCATCCAACTCTTGTTCTG
DAM967	*dprA*	gtcaggatccAGTTATTTATGAAAATCACAAACTATGAAATCT
DAM968	*dprA*	gcatgtcgacTTAAAATTCAAATTCCGCAAGAACATC
DAM969	*comD*	cagtggatccAAAGAGTAATGGATTTATTTGGATTTG
DAM970	*comD*	gcatctcgagCTTTCATTCAAATTCCCTCTTAAATCTA
DAM971	*comE*	gtacggatccGAATGAAAGTTTTAATTTTAGAAGATGTTATTG
DAM972	*comE*	gcatctcgagTCAATCACTTTTGAGATTTTTTCTCTAA
DAM973	*comX*	gtacggatccAGGGGAAAATTATGATTAAAGAATTGTAT
DAM974	*comX*	gcatctcgagCTAATGGGTACGGATAGTAAACTC
DAM975	*comW*	gtcaggatccTTATGTTACAAAAAATTTATGAGCAGATG
DAM976	*comW*	gcatctcgagTACTAAAATTACCTCAACAAGAAATAAAC
DAM989	*recA*	gactggatccGAATGGCGAAAAAACCAAAAAAATTA
DAM990	*recA*	gtcactcgagCAGCTTATTCTTCAATTTCGATTTCA
DAM1047	*Up-lacZ*	AGCAGCCTGCGGCTAGTTTC

DAM1048	*lacZ*	GCGGATTGACCGTAATGG
DAM1049	*Up-lacZ*	GGCATACCAGTCGTTTCAAG
DAM1050	*lacZ*	TCGGTCAGACGATTCATTGG
DAM1051	*lacZ*	GTCGCTACCATTACCAGTTG
DAM1052	*cat*	AATAGCGACGGAGAGTTAGG
DAM1053	*lacZ*	TACCGGATTGATGGTAGTGG
DAM1054	*comA*	CGGACGATAGTGACGTTTC
DAM1055	*comX*	gtac<u>aagctt</u>ATGATTAAAGAATTGTATGAAGAA
DAM1056	*comX*	gcta<u>tctaga</u>ATGGGTACGGATAGTAAAC
DAM1057	*comX*	gtca<u>tctaga</u>GTGGAAGTTACTGACGTAAG
DAM1058	*lacZ*	gcat<u>gaattc</u>CCTGCCCGGTTATTATTAT
DAM1073	*comX*	GGCGTTACCCAACTTAATCG
DAM1074	*lacZ*	gtca<u>tctaga</u>ACGCGATCGGCATAACC
DAM1075	*cat*	gcat<u>gaattc</u>TGTATTTGAGTTTATCACCCTTGTC
DAM1076	*comX*	TATGTATCGTCGCCTTGGTG
DAM1077	*spc*	gtca<u>tctaga</u>CGGGCTGCAGGAATTCG
DAM1078	*spc*	gcat<u>gaattc</u>GGCCGCAGTGTTATCACTC
MSL39	*tet*	AGTTGGCTGGTTACCTTGAATGT
PL05	pCRT7	CCGCGAAATTAATACGACTC
PL06	pCRT7	AAGCTTCGAATTGCCCTTGT
PL08	*comX*	GACCAAGAAGGCATGCTCTG
PL12	pCRT7	CCCTCAAGACCCGTTTAGA
PL82	*aga*	cgc<u>ggatc</u>CTGACTTACTAATGGGTACG
PL103	*spc*	AGCAGTTCGTAGTTATCTTGGAGAG
PL104	*spc*	CAATGGTTCAGATACGACGACTAAA
TTM01	*cps3*	at<u>catga</u>CCTCCCTCGTATTGT
TTM02	*cps3*	cgc<u>ggatcc</u>TTAATAGTGGGAATTTG

a) lower case, 5' extension; b) underscore, restriction site

fragment was purified with QIAquick PCR purification kit to get rid of the cut-off small nucleotides. If plasmid DNA was digested, the digested target sequence sometimes needed to be separated by electrophoresis assay in 1% agarose gel and purified with QIAquick Gel Extracting kit.

2.5.3 DNA Ligation

The concentrations of the digested DNA fragments were measured to estimate the empirical molar factor by dividing the concentration with the size of the DNA. For the typical three-fragment ligation, the molar ratio of the three fragments was used as 1:1:1. For the ligation of vector and insert, the molar ration of vector to insert is 1:3 or 1:5. T4 DNA ligase (Fermentas) was used to ligate DNA fragments with matching overhangs. In a volume of 50μl reaction, 100~300ng of total amount of DNA fragments was mixed with 1~2 μl T4 DNA ligase. The reaction was kept at room temperature for one hour and then incubated at 16ºC overnight.

2.6 Transformation

2.6.1 *S. pneumoniae* transformation

Transformation of pneumococcal cells was started by growing 10-ml culture at 37°C in complete CAT medium with 10 mM HCl to inhibit auto-induction of competence. When the cell density reached 0.03, the culture was further supplemented with 0.04% bovine serum albumin (BSA), 0.5 mM $CaCl_2$, and 250 ng/ml CSP, as described previously (Lee and Morrison, 1999). Cells were then exposed to DNA (~10 ng/ml) at 37°C for 60 minutes before plating in CAT broth agar with appropriate drugs. Serial dilutions of transformed cultures were prepared for plating, ranging from 10^{-1} to 10^{-5}. Four-layer strategy was employed to plate diluted cultures: the first layer was 3 ml complete CAT broth agar; the second layer was the mixture of 1.5 ml diluted cultures with 1.5 ml melted CAT broth agar (kept at 52°C before use); the third layer was 3 ml complete CAT broth agar; then the antibiotics for selecting transformants was added to the fourth layer (top layer) at 4 times the relevant final concentration (Morrison et al.,

1983). The plates were incubated at 37°C for at least 20 hours and the colonies of the transformants were picked. The results were reported as transformants per ml of culture. If genomic DNA was used as donor DNA, the transformation efficiency was about 10^9 cfu/µg; If synthetic linear DNA was used as donor, the transformation efficiency was about 10^7 cfu/µg.

2.6.2 *E. coli* transformation

Electroporation was employed to transform *E. coli* cells. To prepare electro-transformable *E. coli* cells, a single colony was picked and grown in 10ml LB at 37°C overnight with vigorous shaking. The next day, the 10-ml culture was inoculated into 1 L LB broth medium and incubated at 37°C with shaking to OD_{600}=0.5~1.0. The whole culture was chilled on ice and centrifuged at 5000 rpm for 10 minutes at 4°C to harvest the cells. The pellet was washed twice by re-suspending with 800 ml cold sterile H_2O and spun down again. The pellet was resuspended with 20 ml cold sterile 10% glycerol and centrifuged at 4000 rpm for 10 minutes at 4°C. The supernatant was removed again and the pellet was resuspended with 5 ml cold sterile 10% glycerol. Sterile 60 µl aliquots were made into 0.6 ml micro-centrifuge tubes, frozen on dry ice immediately and stored at -80°C up to 6 months.

To transform *E. coli* cells with plasmid DNA, the frozen electro-transformable cell and the electroporation cuvette were put on ice for at least 10 minutes. Less than 3 µl plasmid DNA (~100ng) was taken and mixed with the thawed *E. coli* cell. The mixture was then transferred to the chilled cuvette (1 mm Gap, Fisherbrand) and electroporated with voltage 2.0 kV, resistance 200 Ω and capacitance 25 µF. Then the electroporated cells were transferred to 1 ml LB broth medium and grown at 37°C for one hour with mild shaking. The cell culture was spun down and the pellet was resuspended with 100 µl sterile H_2O. The cell was plated onto two LB plus 100 ng/ml ampicillin agar plates by making approximately 1/10 dilutions: pipetting and spreading 10 µl of the cells onto one plate and 90µl onto another plate. The plates were incubated at 37°C for at least 12 hours and the colonies of the

transformants were picked. Based on the sizes of plasmids I used in this study, 5~8 kb, the average transformation efficiency in my experiment was about 10^7 cfu/μg.

2.6.3 Yeast transformation

The protocol used to transform yeast cells was modified from Gietz et al. (2002). Transformation of yeast haploids was started by growing strains in YPD or SD selective media overnight at 26°C with vigorous shaking. Next day, the cells were spun down, at least 1 ml cells for one transformation. The pellet was resuspended in 0.5 ml PEGLET buffer (40% PEG, 0.1 M LiAC, 10 mM EDTA, 0.1 M Tris-HCl, pH7.5). About 100μg ssDNA (salmon sperm DNA, Sigma) plus 0.5~1 μg plasmid DNA were added to the resuspended cells. The cell culture was seated on the bench at RT overnight. Next day, the cells were spun down again. The cell pellet was resuspended with 100 μl sterile H_2O, and plated onto two selective plates by making approximately 1/10 dilutions: pipetting and spreading 10μl of the cell onto one plate and 90 μl onto another plate. The plates were incubated at 26°C for at least 3 days and the colonies of the transformants were picked. The transformation efficiency for shuttle plasmid pGBDUC2, 6.0 kb, was about 10^3cfu/μg; while for pACT2, 8.1kb, the transformation efficiency was about10^2 cfu/μg.

2.7 Comptence analysis

2.7.1 Competence kinetics assay

To study transformation directly, 1 ml samples were withdrawn from the culture at different times (before and after CSP induction) and mixed with 25 ng of CP1500 chromosomal DNA for continued incubation. After 3.5 minutes, 0.1 ml of the mixture was added to 1.5 ml of CAT with 10 μg/ml DNase I, incubated at 37ºC for another one hour. After that, the 1.5-ml cultures were mixed with 1.5 ml of melted CAT agar, and poured into plates as the second layer of the four-layer plating strategy,

described above (2.6.1). After applying the selective drug onto the fourth layer, plates were incubated at 37°C overnight. Next day, colonies were counted and the data were plotted against the times.

2.7.2 Monitoring late gene expression by β-galactosidase assay

To study transformation indirectly, a *lacZ* reporter gene was placed in *ssbB* operon under the control of ComX to monitor late gene expression. Cultures grown in complete CAT broth plus 10 mM HCl at 37°C were treated with CSP or raffinose at certain cell density. 0.4-ml cultures were sampled at different times after CSP induction (0, 10, 20, 30, 40, 60, 80 minutes). Cells were lysed by mixing with 0.1 ml 5x Z buffer (0.06 M Na_2HPO_4, 0.04 M Na_2HPO_4, 0.01 M KCl, 0.001 M $MgSO_4$, 0.1% Triton X-100, and 0.05 M β-mercaptoethanol) and incubating for 10 minutes at 37°C. 150 µl of cell lysates were mixed with the substrate of 50 µl ONPG (*o*-nitrophenyl-*β*-D-galactopyranoside, 4 mg/ml in 0.1 M phosphate buffer) (Sambrook et al., 2001). Each sample was tripled and the corresponding enzyme level was averaged from them. The activity is expressed in Miller Units with respect to the OD_{550} of the culture when CSP or raffinose was added as shown in the following formula:

$$M.U. = \text{absorption kinetics at } OD_{420nm} / OD_{550nm}$$

where absorption kinetics at OD_{420nm} is the slope of absorptions at OD_{420nm}, and OD_{550nm} reflects the cell density of the culture.

2.8 Electrophoresis assay

2.8.1 Agarose gel electrophoresis

Agarose gels were used for DNA electrophoresis to analyze DNA digestion fragments or PCR fragments. Different concentrations of agarose gel were made in TAE buffer (0.04 M Tris-acetate and 1 mM EDTA, pH8.0) for the analysis of DNA fragments with different sizes. Usually, 0.8% gel was used for examining the chromosome or large plasmids (10 kb or bigger); 1.0% gel was used for detection of smaller plasmids (less than 10 kb); and for fragments of several hundred base pairs, 0.5% gel was

poured. The DNA loading buffer used was type III buffer (6x stock contained 0.25% bromphenol blue, 0.25% xylene cyanol, and 30% glycerol). Gels were run in TAE buffer at 120V till the bromphenol blue approaching the edge. After the running was over, the agarose gels were stained in distilled water containing 0.5 mg/ml final concentration of ethidium bromide (EtBr) for 5 minutes, and destained in distilled water for another 15 minutes before DNA fluorescence was photographed on a UV-transilluminator.

2.8.2 SDS-polyacrylamide gel electrophoresis

Proteins were analyzed by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) using the Bio-Rad protein Mini electrophoresis system. The gel was discontinuous with a lower resolving gel and an upper stacking gel (Sambrook et al., 1989). The resolving gel was poured first, containing 0.375 M Tris-HCl buffer (pH8.8), 15% acrylamide from the stock 30% solution (the ratio of acrylamide and bis-acrylamide is 30:0.8), 0.1% SDS (sodium dodecyl sulfate), 0.05% ammonium persulfate, and 0.05% TEMED. After the separating gel polymerized, a stacking gel solution was poured on the top, containing 0.125 M Tris-HCl buffer (pH6.8), 5% acrylamide/bis-acrylamide, 0.1% SDS, 0.05% ammonium persulfate, and 0.1% TEMED. A comb was placed into the stacking solution to form wells during polymerization. The comb was removed once the stacking gel was solidified. Protein samples were mixed with equal volumes of 2x loading buffer (100 mM Tris-HCl pH6.8, 4% SDS, 0.2% bromophenol blue, 20% glycerol) and denatured at 95ºC for 5 minutes. After the proteins samples were loaded into the wells, electrophoresis was carried out at 53V in stacking gel and at 108V in resolving gel. Polyacrylamide gels for protein analysis were fixed and stained with 0.8% Coomassie Brilliant Blue R-250 in 10% methanol /30% glacial acetic acid for an hour with gentle agitating on a rotary shaker, and distained in 10% methanol /5% acetic acid until prominent blue protein bands against clear background were obtained. Gels were then dried and scanned for images.

2.8.3 Western blot

Beside Coomassie staining, the SDS-PAGE gels could be probed with westernblot if the antibodies were available. To do that, the gel was removed from the electrophoresis apparatus and washed three times with de-ionized H_2O, 100 ml for 5 minutes each time. Meanwhile, the PVDF membrane was briefly washed with 100% methanol for 30 seconds and washed with H_2O twice for 2 minutes each time. Then, the gel, PVDF membrane, fibers and papers are immerged into transfer butter (25 mM Tris, 192 mM glycine, 10% methanol) and equilibrated for 15 minutes. After equilibrium, the transfer package was assembled in this order: white side, fiber, paper, PVDF, gel, paper, fiber and black side. After electro-transferring for 2 hours at 36 V at 4°C, the membrane which contained transferred proteins was washed twice with 50 ml TBST buffer (20 mM Tris, 137 mM NaCl, 1% Tween-20, adjust pH to 7.6 with HCl). Then, the membrane was blocked with TBST containing 5% powdered nonfat milk overnight at 4˚C.

Next day, after rinsing twice for 2 minutes with 25 ml of TBST, the membrane was probed with 20 ml TBST containing first antibody (anti-ComX or anti-ComW, 1:3000 dilution) and 1% powdered nonfat milk for 2 hours at room temperature. After the incubation, the membrane was washed six times with 50 ml TBST for 5 minutes each time. Then, the membrane was probed again with 20 ml TBST containing secondary antibody (anti-rabbit IgG-horseradish peroxidase, 1:20000 dilution) for one hour at RT. Subsequently, the membrane was washed in the same way as above after the application of first antibody.

While washing the membrane, the detection reagents (ESL Plus, Amersham) were removed from storage at 4ºC and allowed to equilibrate to room temperature on the bench before opening. After 30 minutes of equilibrium, 2 ml of solution A and 50 µl of solution B were mixed. The membrane was taken out from the last wash, drained the excess buffer, and placed onto plastic wrap with protein side

up. The mixed solution A and B was then pipette onto the membrane carefully and incubated at room temperature. After incubation for 5 minutes, the detection reagent was drained off by holding the membrane in forceps and touching the edge against a tissue. Then the drained membrane was placed onto another plastic wrap with protein side down and wrapped carefully by smoothing out any air bubbles. The well wrapped membrane was put into a X-ray film cassette with protein side up and taken to a dark room. In the dark room, films were placed on top of the film and exposed for a range of times, such 30 second, 2 minutes, 5 minutes and 10 minutes. The films were eventually developed in a film processor (SRX-101A, Konica).

2.9 Protein purifications

2.9.1 Purification of N-His$_6$-ComX from *E. coli*

To purify ComX from *E. coli*, a copy of comX gene was fused to His$_6$-tag in plasmid pXPL01 and this plasmid was transformed into *E. coli* strain BL21. 10 ml of *E. coli* strain BL21:: pXPL01 in LB broth plus 100 µg/ml ampicillin was grown over night and inoculated in 1 L LB plus 100 µg/ml ampicillin in a 4-L flask. The culture was incubated at 37ºC with shaking at 220 rpm. At OD600 of 0.6, 1 mM IPTG was added to induce *comX* expression and the culture was moved to 30°C for growing another 4 hours. After that, the cells were chilled and harvested with 7,000xg for 20 minutes at 4ºC. The wet pellet was weighted and stored at -80ºC overnight. Next day, the pellet was thawed on ice and resuspended in lysis buffer (50 mM Tris-HCl pH8.0, 1 mM EDTA, 50 mM NaCl, 5% glycerol, 0.1 mM PMSF, 1 mM DTT, and lysozyme to 0.1 mg/ml) with 10 ml per gram of cell pellet. After sonication (38% amplitude, 24 10-second pulses) on ice, the lysate was centrifuged at 10,000xg for 20 minutes at 4ºC. The pellet was washed by resuspending in TGED buffer (50 mM Tris-HCl (pH8.0), 5% glycerol, 0.1 mM EDTA, 1 mM DTT) plus 2% sodium deoxycholate with 10 ml per gram of cells. The washed pellet with inclusion bodies was spun down at 10,000xg for 20 minutes at 4ºC and the wash was

repeated one more time. The inclusion bodies were dissolved in 5 ml TGED per gram of cell pellet plus 0.5% Sarkosyl by stirring for 1 hour at room temperature. After centrifugation at 12,000xg for 30 minutes at 4ºC, the supernatant was diluted with 1.5 volumes of TGED buffer plus 1 mM PMSF. The diluted supernatant was dialyzed against 1.5 L column buffer (50 mM Tris-HCl pH8.0, 300 mM NaCl, 10% glycerol, 5 mM β-mercaptoethanol, 0.1 mM PMSF) twice at 4°C: one time for 2 hours and another time overnight. The dialyzed supernatant was centrifuge at 12,000xg for 30 minutes at 4°C and was loaded onto a nickel affinity column (Ni-NTA Agarose, Qiagen) at 0.25 ml/minute with a pump. After washed with 10 column volumes of column wash buffer (column buffer plus 40 mM imidazole), the bound protein was eluted out with 10 column volumes of elution buffer (column wash buffer plus 250 mM imidazole). Fractions were collected and analyzed on SDS-PAGE. The fractions containing ComX were combined and dialyzed against 1 L storage buffer (50 mM Tris-HCl pH8.0, 50% glycerol, 250 mM NaCl, 0.1 mM EDTA, 1 mM DTT) for 2 hours and can be stored at -20°C or -80°C. After measuring ComX concentration with Bradford assay, the yield of ComX purification was estimated to be 3.5 g per liter culture.

2.9.2 Purification of N-His$_6$-ComX from *S. pneumoniae*

To purify ComX from *S. pneumoniae*, a His$_6$-tag was fused to the N-terminal of *comX* in CP2057, in which *clpP* was disrupted to stabilize ComX. 1.5 L of compete CAT supplemented broth with 10 mM HCl was inoculated with frozen CP2057 stock and incubated at 37°C. 10-ml culture was transferred to a 18 mm by 150 mm tube for monitoring optical density. When the cell density reached OD$_{550}$ of 0.1, the culture was induced with 0.04% BSA, 0.5 mM CaCl$_2$, 0.04% BSA, and 250 ng/ml CSP. At the 15 minutes after the induction, the cells were chilled rapidly using steel beakers and harvested with 7,000xg for 20 minutes at 4ºC. The wet cell pellet was resuspended in 20 ml rinse buffer (10 mM Tris-HCl pH8.0, 300 mM NaCl, 20% glycerol, and 10 mM MgCl$_2$). The cells were spun down

with 10,000xg for 10 minutes at 4°C and stored at -80°C. Next day, the pellet was thawed on ice and resuspended in 10 ml lysis buffer (rinse buffer with 1 mM PMSF, 0.1% Triton X-100, 5 µg/ml DNase I, and 5 mM β-mercaptoethanol) and incubated at 37°C for 10 minutes to lyse the cells. The lysate was centrifuged at 13,000xg for 40 minutes at 4°C. The supernatant was passed through a 0.22 µm syringe filter and loaded onto a nickel affinity column at 0.25 ml/minute. After washing with 10 column volumes of wash buffer (lysis buffer plus 20 mM imidazole), bound protein was eluted with 4 column volumes of elution buffer (lysis buffer with 200 mM imidazole). Fractions were collected and analyzed on SDS-PAGE. The fractions containing ComX were combined and dialyzed against 1L storage buffer (50 mM Tris-HCl pH8.0, 50% glycerol, 250 mM NaCl, 0.1 mM EDTA, 1 mM DTT) for 2 hours and stored at -20°C or -80°C.

2.9.3 Purification of MBP-ComW from *E. coli*

To improved solubility, ComW was purified by using pMAL^{TM} protein fusion system (New England Biolabs), in which a copy of comW was fused to the C-terminal of MBP in expression plasmid pAP01. A 10-ml culture of *E. coli* strain BL21::pAP01 in LB broth supplemented with 2mg/ml glucose plus 100 µg/ml ampicillin was grown over night and inoculated into 1 L LB broth media supplemented with 2 mg/ml glucose plus 100 µg/ml ampicillin next day. The culture was grown at 37°C with shaking at 220 rpm. 10-ml culture was taken out to grow separately in a 18 mm by 150 mm tube for monitoring optical density. When the cell density reached $OD_{600}= 0.6$, the culture was induced with IPTG to 0.3 mM and incubated at 37°C for three additional hours. After that, the cells were chilled and harvested with 7,000xg for 20 minutes at 4°C. The weighed wet cell pellet was stored at -80°C overnight. Next day, the pellet was thawed on ice and resuspended in column buffer (20 mM Tris-HCl pH7.4, 1 mM EDTA, 200 mM NaCl) at 10 ml per gram of cell pellet. After sonication on ice, the lysate was centrifuged at 9,000xg for 20 minutes at 4°C. The supernatant was loaded at 0.25 ml/minute onto an

amylose resin column (New England Biolabs) with a bed volume of 1 ml per gm of cells. After washing

with 12 column volumes of column buffer, bound protein was eluted with 4 column volumes of elution

buffer (column buffer with 10 mM maltose). Fractions were collected and analyzed on SDS-PAGE. 1 μg

of Factor Xa (New England Biolabs) was used to cut 50 μg of MBP-ComW in 50 μl of cleavage buffer

(20 mM Tris-HCl pH8.0, 100 mM NaCl, 2 mM $CaCl_2$) for 12 hours at room temperature. The cleavage

mixture was load onto a hydroxyapatite column and washed with 80 ml of column buffer to wash away

maltose. The proteins were eluted out with 0.5 M sodium phosphate buffer (0.117 M NaH_2PO_4, 0.383 M

Na_2HPO_4, pH7.2). 2-ml fractions were collected and detected for protein by Bradford assay. The

hydroxyapatite-eluted proteins were loaded onto the amylase column again and eluted with 20 ml

column buffer. ComW should be in the flow-through. After measuring ComW-MBP concentration, the

yield of protein purification was estimated to be 25 g per liter of culture.

2.9.4 Purification of His_{10}-RNAP from *S. pneumoniae*

To purify RNA polymerase, a His10-tag was fused to the β'-subunit in strains CP2026 or

CP2027. 10 ml of compete CAT broth inoculated with strains was grown at room temperature overnight

and re-inoculated next day to 1 L complete CAT medium supplemented with 10 mM HCl. At the first

signs of turbidity during growth at 37ºC, 10 ml was transferred to a 18 mm by 150 mm tube for

monitoring optical density. When the cell density reached about $OD_{550}= 0.1$, it was induced to

competence with 0.04% BSA, 0.5 mM $CaCl_2$, 0.04% BSA, and 250 ng/ml CSP. At 15 minutes after CSP

induction, the cells were chilled rapidly using steel beakers and harvested with 7,000xg for 20 minutes at

4ºC. The wet cell pellet was briefly washed in 10 ml rinse buffer (10 mM Tris-HCl pH8.0, 300 mM

NaCl, 20% glycerol, and 10 mM $MgCl_2$). The cells were spun down again with 10,000xg for 10 minutes

at 4°C and stored at -80ºC overnight. Next day, the pellet was thawed on ice and resuspended in 10 ml

lysis buffer (rinse buffer plus 1 mM PMSF, 0.1% Triton X-100, 5 μg/ml DNase I, and 5 mM β-

mercaptoethanol). After lysis by incubation at 37°C for 10 minutes, the lysate was centrifuged at 14,000xg for 40 minutes at 4ºC. The supernatant was passed through a 0.22 μm syringe filter and loaded onto a nickel affinity column with a bed volume of 2 ml of Ni-NTA agarose at 0.25 ml/minute. After washed with 10 column volumes of wash buffer (lysis buffer plus 30 mM imidazole), bound protein was eluted with 4 column volumes of elution buffer (lysis buffer with 150 mM imidazole). Fractions were collected and analyzed on SDS-PAGE.

2.9.5 Bradford assay for measuring protein concentrations

Concentrations of protein were determined by using the Bio-Rad Protein Assay which is based on the Bradford assay (Bradford, 1976). For reconstitution of standard curve, a serial of BSA protein with concentrations ranging from 1.2~10 μg/ml were prepared. 800 μl of each standard or sample solution was pipetted into a clean, dry plastic cuvette. Protein solutions were analyzed in duplicate or triplicate. 200 μl of dye reagent concentrate was added to each cuvette and vortexed. The solution mixtures were incubated at room temperature for at least 5 minutes. The absorbance of each sample at 595 nm was measured. The concentration of the sample protein was calculated by making reference to the BSA standard curve.

2.10 Protein interactions

2.10.1 Yeast two-hybrid assay

To study protein interaction in vivo, yeast two-hybrid assay was established. In yeast two-hybrid, prey and bait proteins were cloned into pACT2 and pGBDUC2 respectively using PCR followed by restriction enzyme digestion and ligation at BamHI and XhoI, or SalI sites. The two plasmids with corresponding inserts were first transformed into *E. coli* (DH5α) to replicate and purified with QIAprep Miniprep kit. The purified plasmid DNA was then transformed into yeast haploids PNS468 (MATE-a) and PNS752 (MATE-α) respectively. The two transformed haploid cells with different mating types

were selected in SD-Leu or SD-Ura media and mated to make diploids in SD-Leu-Ura agar plate by mixing 3µl of each haploid. The diploids were collected and stored at -80 ºC in 20% glycerol stocks. All diploid cells were grown again in YPD agar plate, including the negative controls harboring plasmids without inserts, or with single insert. After two day's incubation at 26ºC, the diploid cells were frog-replicated onto four test agar plates: SD-Leu-Ura; SD-Leu-Ura-His; SD-Leu-Ura-His+1mM 3AT; SD-Leu-Ura-His+3mM 3AT. The four test plates were then incubated at 26 ºC for 7 days. Pictures were taken every day once there were colonies growing up.

2.10.2 Surface plasmon resonance (SPR)

To explore protein interaction in vitro, SPR was employed with Biacore T100. Two proteins were expressed and purified in vitro, such as through His-tag purification or MBP purification. The smaller protein was immobilized onto the sensor chip as the ligand and the bigger one was used as the analyte in the flow. The amount of ligand bound to the chip was about 50 µg and amount of target put in the flow was about 0.75 µg. CM5 sensor chip was selected for ligand immobilization. In ligand immobilization, the dextran matrix of CM5 chip was first activated by the mixture of 100 µl 0.4 M EDC (1-ethyl-3-(3-dimethylaminopropyl) carbodiimide) and 100 µl 0.1M HNS (N-hydroxysuccinimide), producing reactive succinimide esters. The ligand in acetate buffer with certain pH value (10 mM, pH4.0~6.0) was then passed over the surface and the esters reacted with amino groups to form covalent bonds. The remaining esters were subsequently blocked with 100 µl 1M Ethanolamine, pH8.5. After ligand immobilization, the analyte was diluted into 150 µl HBS-EP running buffer (0.01 M HEPES pH7.4, 0.15 M NaCl, 3 mM EDTA, 0.05% surfactant P20) to make 50 nM concentration. The solution was further diluted 10 times to make 150 µl 5 nM low concentration. Then, the analytes in two concentrations were passed over the chip surface at 25 µl/minute for 2 minutes and the sensogram was monitored.

Chapter III. Construction and Evaluation of A New Self-deleting Cre-*lox-ermAM* Cassette, CHESHIRE, For Marker-less Gene Deletion in *Streptococcus pneumoniae*

3.1 Abstract

Although targeted mutagenesis of *Streptococcus pneumoniae* is readily accomplished with the aid of natural genetic transformation and chimeric donor DNA constructs assembled in vitro, the drug resistance markers often employed for selection of recombinant products can themselves be undesirable by-products of the genetic manipulation. A new cassette carrying the erythromycin-resistance marker *ermAM* is described that can be used as a temporary marker for selection of desired recombinants. The cassette may subsequently be removed at will by virtue of an embedded fucose-regulated Cre recombinase gene and terminal *lox66* and *lox71* Cre recognition sites, with retention of 34 bp from the cassette as an inert residual double-mutant *lox72* site.

3.2 Introduction

Streptococcus pneumoniae (pneumococcus), a Gram-positive (Gm⁺) bacterium which normally inhabits the human upper respiratory tract, is an important opportunistic pathogen that causes a variety of infections and diseases. As a naturally transformable species, it is especially accessible to genetic manipulation (Morrison, 2007) and many of its virulence factors have been identified through intensive genetic analysis. In such genetic analysis, it is sometimes desirable to construct strains with double, triple, or multiple gene disruptions. Especially in the case of multiple disruptions, it is often preferable that such disruptions do not themselves cause accumulation of heterologous proteins in the resulting mutant. While a variety of marker-less gene disruption strategies have been developed for application in pneumococcus (Iannelli et al., 2004; Kloosterman et al., 2006; Standish et al., 2005), they involve either multiple genetic manipulation steps or tedious screening steps.

The Cre/*loxP* strategy for creating marker-less deletions is especially attractive, as it uses the well-characterized, naturally occurring cofactor-independent site-specific recombinase of bacteriophage P1 to delete arbitrary targets delimited by two copies of the 34-bp *loxP* recognition sequence (Ghosh et al., 2002; Leibig et al., 2008). Furthermore, use of carefully chosen single-mutant *loxP* sites can ensure that a residual double-mutant *loxP* site is produced that does not participate in further Cre-mediated recombination (Albert et al., 1995; Lambert et al., 2007). Several implementations of this strategy for use in Gm⁺ bacteria have already been described (Banerjee et al., 2008; Kloosterman et al., 2006; Lambert et al., 2007; Leibig et al., 2008; Pomerantsev et al., 2006; Yan et al., 2008). In practice, two directly repeated *loxP* sites are arranged to flank a selectable marker, which is substituted for the deletion target by use of targeted recombination. Then, upon expression of a Cre recombinase gene, recombination between the *loxP* sites excises the intervening sequence, leaving one residual recombinant *loxP* element in place of the deletion target. Even here, however, current bacterial implementations typically employ a four-step strategy of (I) emplacing the selectable marker and single-

mutant *loxP* sites as a substitute for the intended deletion target, (II) introducing a heterologous *cre* gene into the resulting mutant, (III) allowing expression of *cre* and excision of the selectable marker, and (IV) removing the heterologous *cre* gene from the resulting deletion mutant. To simplify this process for use in pneumococcus while taking advantage of its highly efficient natural transformation system, I sought to combine steps (I) and (II) and obviate step (IV) by creating a new self-deleting *lox/erm/cre/lox* cassette, much as has been implemented in the plant system of Arabidopsis (Hare et al., 2002; Zuo et al., 2001). Because the excision would be irreversible, the level of expression of the *cre* gene in such a cassette could in principle be adjusted to provide stability high enough for steps (I) and (II), but low enough to provide easy recovery of deletions at step (III). I have chosen instead to place *cre* under the control of a native regulated pneumococcal promoter that has a low basal level of expression, but is readily activated. In this note, I describe construction of such a self-deleting cassette, and show that it is stable in glucose medium but is readily excised during growth in the presence of fucose.

3.3 Results

 3.3.1 Construction of Cheshire cassette.

A *loxP/ermAM/P$_{fcsK}$/cre/loxP* cassette, which we propose to name Cheshire, was constructed by using PCR amplification, restriction nuclease digestions, and ligation (Fig. 7). One fragment was amplified from the *ermAM*-t1t2-*P$_{fcsK}$* cassette, in which the *ermAM* erythromycin (Em) resistance determinant from plasmid pAMβ-1 and the *fcsK* promoter (P$_{fcsK}$) from the fucose operon of *S. pneumoniae* are linked to a strong rRNA transcriptional terminator t1t2 to minimize read-through transcription toward P$_{fcsK}$ and *cre* (Ng et al., 2003). The *cre* gene was amplified from the vector pCrePA (Pomerantsev et al., 2006). Mutant *loxP* sites (Albert et al., 1995) were assembled at the ends of the cassette by designing half-*loxP* sites into the amplicons containing *ermAM*-t1t2-P$_{fcsK}$ and *cre,* while the other half of each *loxP* site was provided in the flanking fragments used to target recombination into the
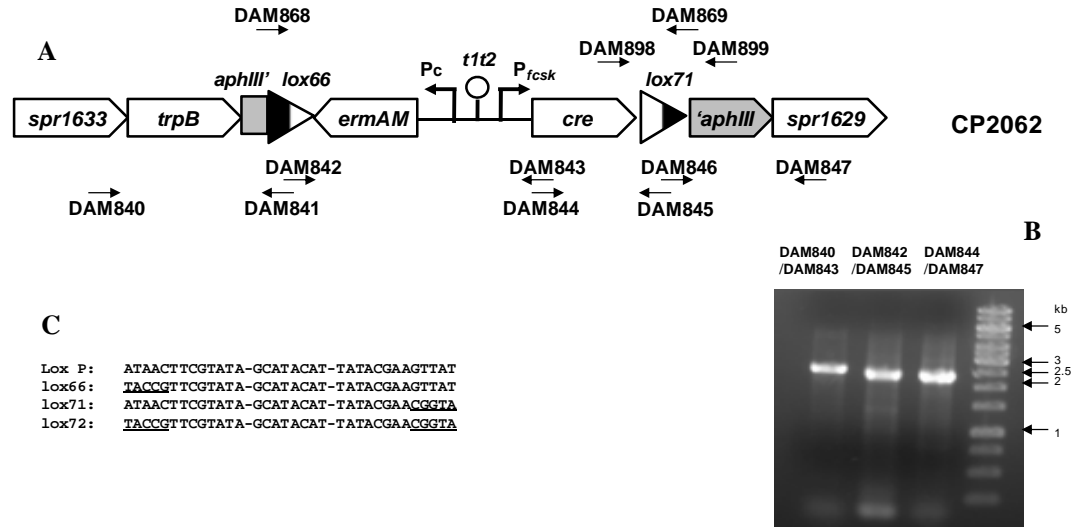
Figure 7 A. Construction of lox/erm/cre/lox cassette as an insertion into a kanamycin resistance gene of *S. pneumoniae*. An *ermAM-t1t2-P$_{fcsK}$* fragment was amplified with the primer pair DAM842/843 using *erm-P$_{fcsK}$* DNA as template. A *cre* fragment was amplified with primer pair DAM844/845 from the pCrePA plasmid. Upstream and downstream fragments were amplified with primer pairs DAM840/841 and DAM846/847, respectively, from genomic DNA of CP1334. The four resulting fragments were digested with *Bsa*I and ligated to produce siteless junctions. After transformation into CP1334, the Em^R strain CP2055 was isolated. The assembled locus in CP2055 was transformed into CP2000, to create CP2062. Lollipop, t1t2 terminator. Pc, constitutive promoter inducing ermAM. *lox66* and *lox71*, left-end and right-end single-mutant *loxP* sites. Kan', 'Kan, *aphIII* truncations. Small arrows show positions of indicated primers. Pentagons, ORFs. Triangles, *loxP* sites, filled to indicate mutant 13-bp repeats within *loxP*. Elements not drawn to scale. **B. Proof of the structure of the cassette in the Em^R clone CP2062.** PCR products obtained with primers pairs DAM840/843, DAM842/845, and DAM844/847. Expected sizes of the products were 2.6, 2.4, and 2.3 kb, respectively. In marker lane, bands of sizes 1, 2, 3, and 5 kb are indicated by arrows. **C. Wild type, single mutants and double-mutant *loxP* sites.** Mutant bases within 13-bp inverted repeats are underlined; illustrative (*MluI* and *NcoI*) restriction sites and linker bases are shown in lower case for the case of a casette prepared using primer pair DAM900/901.

genome by natural transformation. The resulting cassette was inserted into a kanamycin (Kan) resistance gene, *aphIII*, that had replaced the genes spr1630 and spr1631 (Peterson et al., 2004), to obtain CP2062, a Kan^S Em^R insertion derivative of CP2000 (Fig. 1). The disruption of *aphIII* was designed so that precise Cre-mediated excision of the cassette would be expected to restore an *aphIII* ORF (*aphIII::lox72*) with a *lox72* element inserted between the second and fourteenth codons (Ala2 and Ile14). Culture media and transformation of competent cells were as described previously (Lee et al., 1999). During construction, the *Bsa*I asymmetric restriction site was included in junction primers to allow formation of ligation junctions free of residual nuclease recognition sites (Table 1). PCR was used to confirm the structure of the construct and the structure of each new junction in the resulting cassette was verified by sequencing. Finally, primer pair DAM900/DAM901 was designed to allow applications of the cassette using terminal *MluI* and *NcoI* sites. The sequence of the cassette as amplified using the primer pair is deposited in Genbank as Accession bankit-1147842.

3.3.2 Regulation of promoter P$_{fcsK}$ in fucose

Because fucose-induced expression from P$_{fcsK}$ is reported to increase strongly during approach to stationary phase in *S. pneumoniae* (Chan et al., 2003), I sought a treatment regimen in which a high level of expression from P$_{fcsK}$ could be maintained for an arbitrary period. To characterize the transcriptional response of the fucose promoter (P$_{fcsK}$), I used a P$_{fcsK}$::*lacZ* fusion constructed at the *ssbB* locus (Fig. 8). Two portions of a CP2052 culture were treated with fucose, one at early log phase and one at late log phase, and maintained at respective densities by periodic dilution with fresh fucose medium. Compared to the culture without fucose induction, LacZ activity increased 5~10 fold in the culture maintained at OD 0.04~0.08, but 70~140-fold in the culture held at OD 0.4~0.8 (Fig. 8). This is consistent with the report by Chan et al. (2003), and suggests that the rate of excision could be highest in cultures grown with fucose to high cell densities.
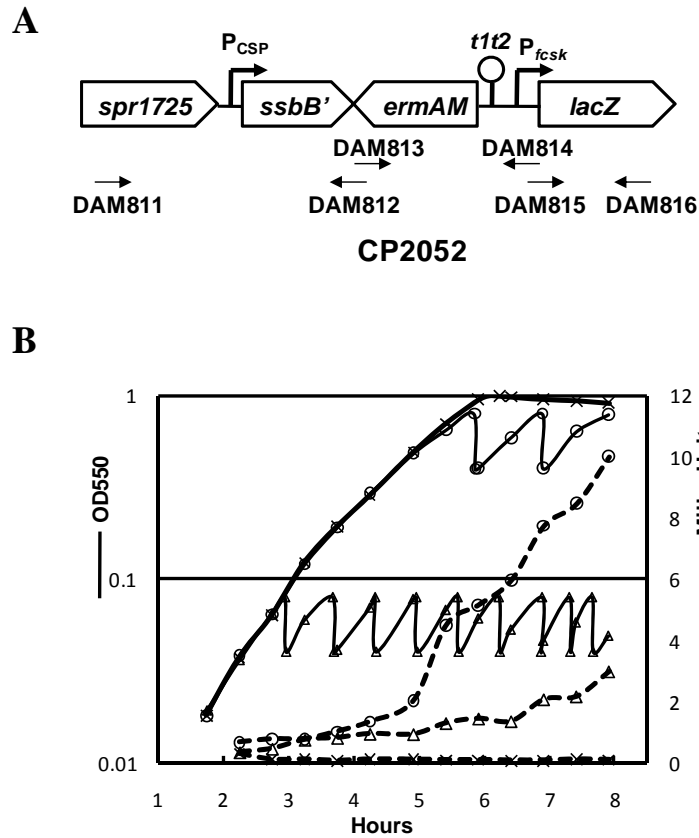
A

B

Figure 8 Effect of fucose on P$_{fcsK}$-driven expression. A. Construction of an ectopic *PfcsK::lacZ* transcriptional fusion in the *S. pneumoniae* genome. A fucose promoter fragment was amplified with primer pair DAM813/814using *ermAM-t1t2-P$_{fcsk}$* DNA as the template. *lacZ* and *ssbB* fragments were amplified with primer pairs DAM811/812 and DAM815/816 from genomic DNA of CPM7. The three fragments were assembled by ligation after digestion at *Eco*RI or *Bam*HI sites included in the junction primers. After transformation into CPM7 with selection for Em^R, the assembled construct was transformed into the CP2000 background to create CP2052. P$_{csp}$, CSP-inducible combox promoter of *ssbB*. **B. Effect of fucose on expression of *P$_{fcsK}$::lacZ* in CP2052.** Three cultures of CP2052 were incubated at 37 C in parallel, with 0.5% fucose added to two cultures at time 0. One fucose-supplemented culture was maintained between OD$_{550}$=0.04 and OD$_{550}$=0.08 by repeated dilution with fucose-CAT broth. The second fucose-supplemented culture was maintained between OD$_{550}$=0.4 and OD$_{550}$=0.8 in the same way. Samples of the cultures taken at the indicated times were assayed for β-gal activity. Solid lines, OD$_{550}$; dashed lines, β-gal activity; ×, no fucose; O, Δ, 0.5% fucose.

3.3.3 Excision of the Cheshire cassette upon treatment with fucose

To characterize the stability and inducible excision efficiency of the new cassette, I examined cultures maintained under conditions for basal or maximally elevated expression of Cre. Two parallel cultures of mutant CP2062 were examined: one culture was incubated without fucose, while another was incubated with fucose from OD_{550}= 0.001 and then maintained by repeated dilution at late log phase between OD_{550} of 0.4 and 0.8 (Fig. 9). As expected, the two cultures grew at the same rates up to OD_{550}=0.8. Samples were taken from both cultures at OD_{550}=0.8, as well as 1 or 2 hours after the first dilution. Each sample was plated in non-selective CAT agar for single-colony isolation. Fifteen colonies obtained from each sample were individually used to inoculate 2 ml of CAT broth. On reaching visible turbity (OD_{550}=0.02), a sample of each subculture was diluted 1:100 into three tubes containing CAT of CAT supplemented with 0.05 µg/ml Em or 200 µg/ml Kan. After 5 hours incubation at 37ºC, all cultures were visibly turbid, but some of the Em cultures of clones from the fucose-treated culture failed to grow, in a proportion that was higher for longer periods of exposure to fucose (Fig. 9). PCR amplification of samples of the CAT cultures using primers to detect the presence of a spr1631-*aphIII* junction fragment revealed that the cassette was present only in subclones that exhibited growth in CAT + Em; all Em[S] clones had lost the cassette (Fig. 9). The construct was quite stable during growth without fucose, as 100% of the subclones from the culture were Em[R] and had maintained the Cheshire cassette during 8 hours incubation (~12 generations). In contrast, the percentage of erythromycin resistant colonies decreased rapidly during exposure to fucose. On reaching OD_{550}=0.8 in fucose medium, 40% of cfu were already free of the *ermAM* cassette, while after two further hours of growth with fucose at high cell density, an additional 40% of the cfu had become erythromycin sensitive and had lost the *cre-aphIII* junction fragment, indicating excision of the cassette from the genome. To determine whether the Em[S] cultures carried the predicted *lox72* recombinant *loxP* site, two of the resulting Em[S] progeny were
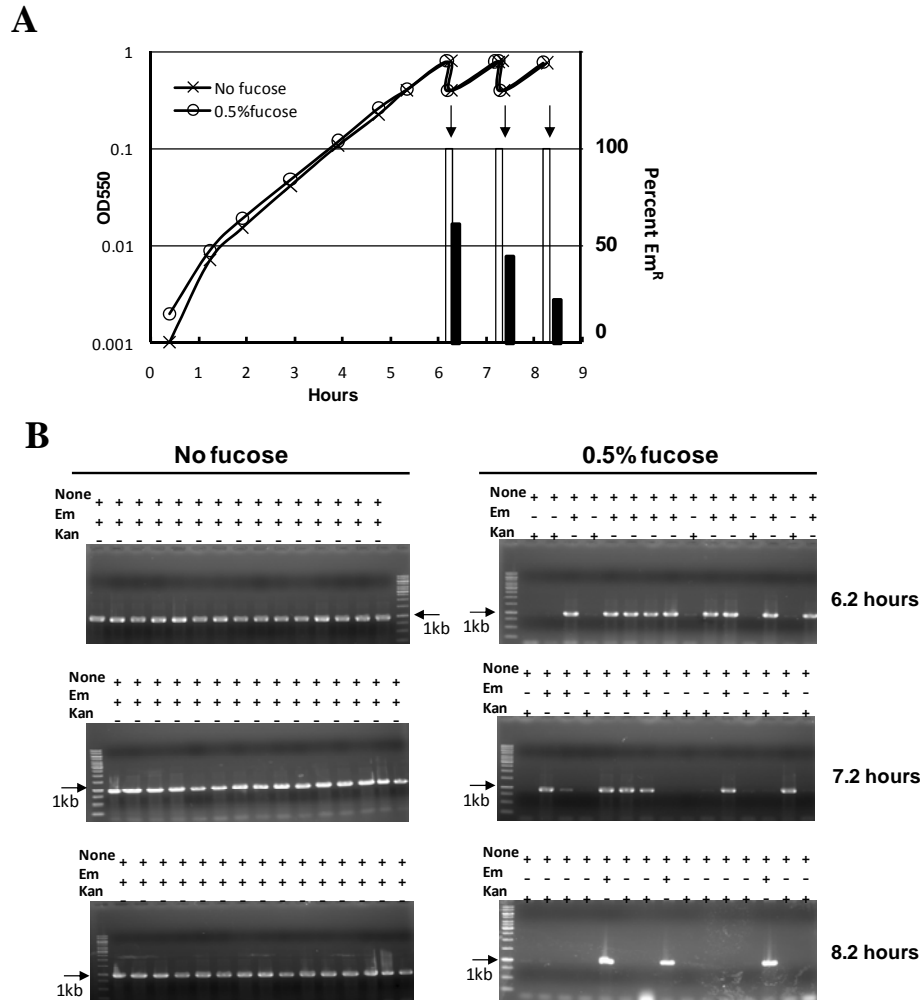
A

B

Figure 9 Fucose-driven excision of Cheshire. A. Growth and Em^R composition of cultures of CP2062. Parallel cultures were inoculated in CAT broth with or without 0.5% fucose at OD_{550}=0.001. Upon reaching OD_{550}=0.8, both were maintained by repeated dilutions between OD_{550}=0.4 to OD_{550}=0.8. Samples taken on reaching OD_{550}=0.8 and one and two hours later, as indicated by the arrows, were diluted 10^6–fold and plated in CAT agar for isolation of single colonies. The percentages of Em^S clones were estimated from analysis of 15 isolated colonies from each sample, as shown in Panel B. Open bars, the percentage of Em^R clones from the culture without fucose; filled bars, Em^R from the fucose-treated culture. Results shown are representative of three similar experiments. **B. Clonal analysis of cultures after fucose treatment.** 15 colonies obtained from each sample described in Panel A were analyzed as described in the text. Loss of the 1-kb PCR product indicates the excision of the cassette. Left images, results for the subclones from the culture grown without fucose; right images, results for those from 0.5%fucose-treated culture. +, growth in the indicated medium. -, no growth. Arrows, position of a 1-kb reference DNA band.

characterized by amplifying and sequencing the spr1632/*aphIII* region. In both cases, the results revealed that the entire Cheshire cassette was indeed precisely replaced by a single *lox72* site, which provided the evidence suggesting the deletion tool worked successfully.

The *aphIII::lox72* gene is predicted to encode a protein with 11 amino acid residues inserted between Ala2 and Ile14 of the native AphIII enzyme. None of the 90 subcultures in CAT+Kan broth were turbid after 6 hours incubation at 37ºC, nor did diluted samples from colonies in agar with 800 μg/ml Kan, indicating that the extended AphIII protein is inactive. However, by 20 hours' incubation at 37ºC, all CAT+Kan cultures of the Em^S subclones had become turbid (Fig. 9). Sequencing of the *aphIII* gene in pure subclones of two of these cultures revealed an intragenic suppressor mutation that apparently had restored sufficient AphIII activity to confer Kan resistance. Both had a G5A substitution within the *lox72* element, which causes an R4H substitution in its predicted protein product. As all Em^S revertants examined were similarly revertible (Fig. 9), but none of the Em^R clones were, this pattern further confirms the absence of the *ermAM* insert, and restoration of the *aphIII* ORF. I conclude that the Cheshire cassette is quite stable in the pneumococcal genome, but is readily excised during a single cycle of culture with supplemental fucose.

3.3.4 Application of the Cheshire cassette as a temporary marker

Cheshire is designed to be used as a temporary Em^R marker for effecting various genetic rearrangements in pneumococcus. To delete a target sequence, for example, two sequences flanking the deletion target can be amplified from the genome and linked to the cassette to create a chimeric donor for transformation of competent cells (Fig. 10). After integration of the cassette in place of the target gene through homologous double cross-over recombination, and selection with erythromycin, the resulting Em^R mutant can be grown without erythromycin but in the presence of fucose for expression of the *cre* gene. Cre-catalyzed excisive recombination between the two single-mutant *loxP* sites, *lox66* and
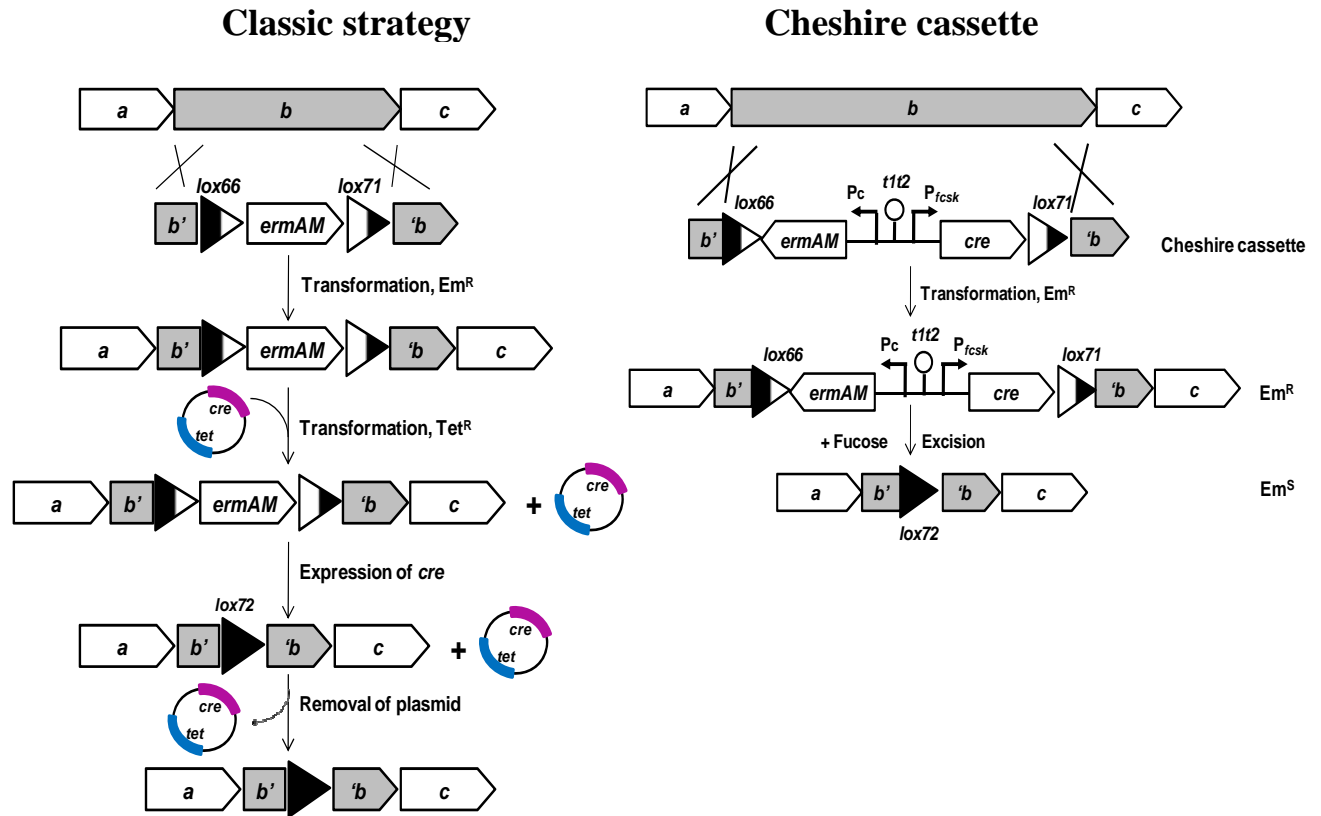
Figure 10. Applications of the cre-lox system to create gene deletions. Left, classic strategy of using cre-lox system to make gene deletion involves four steps: 1) emplacing the selectable marker and single-mutant *loxP* sites as a substitute for the intended deletion target through transformation, 2) introducing a heterologous *cre* gene into the resulting mutant through another transformation, 3) allowing expression of *cre* and excision of the selectable marker, and 4) removing the heterologous *cre* gene from the resulting deletion mutant. **Right**, strategy of using Cheshire cassette to make gene deletion in *S. pneumoniae* just needs two steps: 1) emplacing the Cheshire cassettes as a substitute for the intended deletion target, 2) inducing the expression of *cre* gene by fucose to pop out the cassette. a and c, the corresponding upstream and downstream genes. b', 'b, residual portions of gene b. Cassette elements indicated as in Fig. 7A.

lox71, leaves a residual double-mutant *loxP* site known as *lox72*, reminiscent of the residual smile of the Cheshire cat (Dodgson, Charles Lutwidge, 1866). If excision is efficient, minimal screening will readily identify Em^S excision products. As the *lox72* 'scar' carries only a single stop codon in each strand, it permits the design of in-frame deletions and, as it is not recognized by Cre recombinase (Kloosterman et al., 2006; Zhang et al., 2002), further *loxP*/Cre manipulations may be carried out in the new strain without the danger of mis-targeting of Cre.

3.4 Further discussion.

Marker-less gene deletion methods are valuable in the construction of null mutants or other targeted rearrangements, especially in making multiple modifications in a single strain. In *S. pneumoniae*, several methods had been used to achieve this goal, but each is somewhat more complex than use of the present construct. Iannelli *et al.* (2004) and Standish *et al.* (2005), for example have used a method for introducing unmarked mutation into the genome of *S. pneumoniae* that employs SOEing PCR to construct markerless donor DNA and screening for integration of the deletion by numerous PCR assays. Although successful, this approach entails considerable lab work and depends on achieving a high level of competence. A drop-in/pop-out method for transferring arbitrary unmarked mutations into the pneumococcal genome was developed by Kloosterman et al. (2006), using a blue/white colony discrimination step to facilitate the screening for rare spontaneous pop-out excision clones, but as the mutation must first be assembled in a shuttle plasmid, it requires the added step of plasmid cloning of the mutant construct. Another method, developed by C.K. Sung, et al. (2001), using the JANUS cassette, which contains the recessive streptomycin resistance gene *rpsL*, entails less labor, as it utilizes direct selection at both steps of transformation, but still requires two steps of transformation and a preexisting Sm^R mutation in the genome, and can be complicated by a significant background rate of Sm^R gene conversion events.

The Cre/*lox* system has enabled efficient routes to making marker-less deletions for broad applications in a large variety of species, including plants, the mouse, yeasts and bacteria (Albert et al., 1995; Pomerantsev et al., 2006; Sauer et al., 1987; St-Onge et al., 1996). In this paper, I describe a new *loxP/cre/loxP* cassette suitable for use in pneumococcus that simplifies the process by saving two steps commonly used in bacterial Cre/*loxP* strategies: introduction of a Cre plasmid by transformation and the removal of the Cre plasmid after completion of *lox* recombination (Fig. 10). However, the rate-limiting step of using this tool would be transforming the cassette into the genome, which depends on the sizes of flanking homologous DNA. It has been shown that the transformation efficiency varied over a 6-log range as the length of homology varied between 100bp to 2000bp (Lau, et al., 2001).

The new cassette itself carries the tools for its own removal: the Cre recombinase gene and a promoter that places the *cre* gene expression under control of the fucose-responsive regulator, FcsR. Homology searches among 1354 sequenced bacterial genomes revealed that the fucose operon is widely conserved in *S. pneumoniae*, but is generally absent among other streptococci (except *S. suis*), and that the putative fucose regulator gene *fcsR* (spr1974 in strain R6) has a similar distribution. Thus, it is to be expected that this cassette will be widely useful in *S. pneumoniae*, but rarely in other hosts. However, the simple steps of replacing the fucose promoter by other inducible promoters, with suitably low basal expression according to the specific species of interest, or inclusion of *fcsR* in the cassette could easily adapt this *lox*/*ermAM*/*cre*/*lox* design to other bacteria. Due to its markerless property in gene deletion and simplicity in application, the Cheshire cassette has been shared with more than 30 laboratories around the world and, at least, two successful applications have been published so far (Fontaine et al., 2010; Tran et al., 2011). More contributions are expected in the future as to genetic research in pneumococcus.

Chapter IV. Exit From Competence for Genetic Transformation in *Streptococcus pneumoniae* Is

Regulated at Multiple Levels

4.1 Abstract

Development of natural competence for genetic transformation in *S. pneumoniae* entails two sets of genes induced by the pheromone CSP (**C**ompetence-**S**timulating-**P**eptide): the early genes and the late genes. Early gene expression depends on ComE, a response regulator that is activated in response to CSP. Subsequently, an early gene product, ComX, activates expression of late genes, which eventually makes the cell competent. Expression of both early and late genes is transient, with late gene expression lagging ~5 minutes behind that of early genes. After competence develops, CSP-dependent gene expression is rapidly shut off by a mechanism that is related to ComX-dependent repressor(s). I screened individual late gene mutants to investigate further the roles in competence termination. Among the 20 genes tested, Δ*dprA* displayed a prolonged late gene expression pattern, whereas mutants lacking *cbpD, cibABC, cglEFG, coiA, ssbB, celAB, cclA, cglABCD, cflAB, or radA*, exhibited a wild-type temporal expression pattern. Thus, no other gene than *dprA* was found to be involved in shutoff by this test. Combined with previous lab work showing that DprA limits the amounts of ComX and another early gene product ComW, I concluded that DprA controls their amounts by inhibiting early gene expression rather than by promoting the degradation of ComX and ComW. To ask what target DprA might work on in turning off early gene expression, yeast two-hybrid assay was employed to investigate protein interactions between DprA and ComD, DprA and ComE. My results suggest that DprA could interact with ComE, but not with ComD, ComX, and ComW. Therefore, I further hypothesize that DprA shuts off early gene expression and competence by a direct interaction with ComE.

4.2 Introduction

In laboratory cultures of *S. pneumoniae*, an outburst of competence occurs during the mid-log phase, emerging and shutting off rapidly, within about 30 minutes. During the period of maximal competence, also termed the X state (Claverys et al., 2006), transcription is dominated by an excess of the alternative sigma factor, ComX, which is otherwise entirely absent from the cell. To escape from this state, which is maintained by the positive autocatalytic circuit comprising products of the genes, *comA, comB, comC, comD,* and *comE,* key connections in the circuit must be interrupted decisively. A dramatic temporal pattern of mRNA accumulation and loss in response to an acute dose of the CSP signal attests to the coordination and strength of these effects (Alloing et al., 1998; Peterson et al, 2004; Dagkessamanskaia et al., 2004). A brief period of early gene expression is followed by a brief period of late gene expression and a somewhat longer period of competence reflecting the activities of the accumulated late gene products.

The lag in expression of the late genes is explained by the role of the early gene product ComX as an alternative sigma factor driving expression of late genes from specific non-canonical promoters. However, as rapidly as this chain of responses to CSP is established, its effects are nearly as rapidly reversed (Luo and Morrison, 2003; Peterson et al., 2000; Peterson et al., 2004; Alloing et al., 1998), leading to the remarkably transient nature of competence. Early gene mRNA quickly accumulates, increasing at least 100-fold between 2 and 10 minutes after a sudden increase in CSP level, but these messages then disappear just as quickly, leaving less than 10% of maximal levels in the cell by 15 minutes. As no specifically targeted anti-mRNA mechanism is known, this suggests that ComE-directed transcription stops abruptly by 10 minutes. As the ComE protein itself is stable (Ween et al., 1999), a prime candidate for the cause of this change is a change in the the phosphorylation state of ComE. Transcripts of late genes follow a delayed but similar temporal course, peaking at 13 minutes and largely

disappearing by 17 minutes, implying as well that ComX activity is largely dissipated by 13 minutes. Although labile, ComX protein is present well after 20 minutes, suggesting that some inhibition of ComX activity occurs before it physically decays. A further indication of an antagonist of ComX activity was recently provided: while mutation of ClpP stabilized ComX, it failed to prolong its transcriptional activity, detected either as transformation or of transcription of late genes (Piotrowski et al., 2009). Thus, there appear to be at least two targets of competence shutoff mechanisms, the activity of ComEP (which might be inactivated by a phosphatase) and the activity of ComX.

Martin et al. (2000) reported that the ComE R120S mutant exhibits a much delayed shutoff of competence, with transformation continuing past 90 minutes, much longer than the 20-minute period of competence in wild type cultures. This behavior indicates that ComE activity is an important determinant of the exit from competence and that cells in a competent culture can support an extended period of transformation. Indeed, Claverys and Håvarstein (2002) proposed that ComE itself acts to shut off its own activity when accumulated to a high enough level. It has been thought that exit from competence might be a two-step process involving, first, dephosphorylation or other inactivation of the stable regulator, ComE, followed by degradation of the unstable proteins, ComX and ComW (Claverys and Håvarstein, 2002; Luo and Morrison, 2003). More recently the case for the importance of the second, proteolytic, step was clouded when it was observed that competence still shuts off in a strain where ComX and ComW are stabilized by inactivation of cognate proteases (Piotrowski, Luo, and Morrison, 2009). Furthermore, competence induced by ectopic expression of *comX* and *comW*, without participation of other early genes, also follows a course leading to rapid shutoff (Luo, Li, and Morrison, 2004). These results together indicate that there is at least one factor shutting off competence by targeting ComX directly, independent of proteolysis of ComX and independent of any regulatory effects on expression of early genes.

Additional clues to the mechanism by which competence is terminated are provided by the phenotypes of *comX* and *dprA* mutants. While *comX* mutants do not transform, or become competent by the criterion of expression of late genes, they do respond to CSP, with over-expression of ComX (Lee and Morrison, 1999; Peterson et al., 2000) and other early genes (Peterson et al., 2004). Remarkably, the temporal pattern of this response is unusual: it is not rapidly reversed, as in wild type; instead the induced early gene expression continues for generations (Lee and Morrison, 1999). This simple result immediately suggests that one or more late gene products may be important for the reversal of the CSP response, although another logical possibility is that ComX itself acts as such a repressor. A remarkably similar regulatory phenotype was described for mutants defective in the late gene, *dprA*. Bergé reported that the *dprA* mutation blocked transformation (later explained by roles of DprA in stabilization or recombination of donor DNA fragments), but also caused an exacerbated (~60 minutes) growth arrest upon CSP treatment, while simultaneously permitting continued expression of the *comCDE* early competence operon (Bergé, Univ. Paul Sabatier PhD thesis, 2002).

To search more broadly for late gene(s) implicated in reversal of the response to CSP, Peterson et al. (2004) examined the competence kinetics of many mutants defective in genes that were induced in competent cells but not required for transformation. None of the transformable mutant strains tested displayed an extended period of competence. The late genes tested in this way were SP021, SP030-1, SP123-6, SP0200, SP0782, SP1088 (*radC*), SP1478-80, SP1980-1, SP2046, SP956, SP1808, and SP2201. *radC* (SP1088) mutants were also reported to have normal shutoff kinetics (Attiaech and Martin 2008). Thus, nearly 20 genes that exhibit ComX-dependent expression but are not required for transformation appear to play no critical role in competence exit. Mirouze et al. (2007) subsequently demonstrated that it was possible to complement the regulatory defect of a *comX* mutant in escape from the CSP response, restoring transient expression of the *comCDE* operon, by ectopic *comX*-independent

expression of *dprA,* under control of an early class promoter. They proposed that the relevant activity of DprA might be either to promote de-phosphorylation of ComE or to block phospho-transfer to ComE by ComD, either of which would be expected to cause a broad effect on all early genes in addition to the observed effect on the *comCDE* operon.

Although DprA thus seemed necessary for shutoff of the (early) *comCDE* operon and sufficient for ComX-independent shutoff of *comCDE* transcription, its regulatory target remained uncertain, and it was unknown whether additional late genes might contribute as well to exit from the X state. To identify possible additional pieces of this puzzle, I looked further among the late genes for products that act to promote the exit from competence. Although the late genes outlined in red in Fig. 4, Chapter 1 (*cclA, celA, cflA, cglA, coiA, radA, recA,* and *ssbB*) are required for transformation and have critical roles in DNA transport, processing, or recombination, they had not been examined for possible roles in competence termination, perhaps because of the defective transformation in such mutants. This impediment could be circumvented with an indirect method, simply by monitoring shutoff of late gene expression instead of decay of transformability *per se* as an indicator of exit from the competent state. Since the late gene expression pattern is itself a reliable indicator of competence kinetics (Bergé et al., 2002), the late gene expression pattern revealed by a simple LacZ assay in a ComX$^+$ background will reflect competence development and persistence, even in non-transformable mutants. Using this strategy, with a *lacZ* transcriptional reporter at the *ssbB* late gene, I examined mutants defective in additional late genes that are required for transformation and found only *dprA* mutant displayed prolonged late gene expression. Further endeavors were also made to investigate DprA's target in controlling early gene expression and competence development.

4.3 Results

4.3.1 Among candidate late competence genes examined, only *dprA* is required for normal shutoff of late gene expression in a wild-type background.

To investigate the roles of additional late genes in competence termination, I monitored the expression pattern of a late gene transcriptional reporter, which is established as a reliable indicator of competence development (Bergé et al , 2002). For this purpose, I constructed a parent strain that contains a *lacZ* reporter at the late gene, *ssbB* and is also *comA* deficient, so as to avoid any potential growth deficiency that might be caused by prolonged competence in a mutant deficient in exit from competence. The late gene expression pattern in response to CSP in the *comA* Cm^R Em^R reporter parent strain was verified as identical to that reported previously (Pestova et al.,1996; Lee and Morrison, 1999; Peterson et al., 2000, 2004).

Mutations in genes with known roles in transformation were crossed into the parent strain. Included were *cibABC* and *cbpD,* which affect fratricide, and *cinA,* linked to the combox promoter required for induction of *recA* by CSP. The internal deletion of each gene or operon was replaced by a Kn^R marker. Altogether, I examined 12 mutants, defective in a total of about 20 late genes: *ΔcbpD, ΔcibABC, ΔssbB, ΔcglEFG, ΔcoiA, ΔdprA, ΔcelAB, ΔcclA, ΔcglABCD, ΔcflAB, ΔradA* and *ΔPc-cinA.* Each mutation's structure was confirmed by PCR, and the border of each deletion is shown in Fig. 11. Transformation rates determined for the new strains were comparable to the values expected from the literature (Table IV).

To determine the effect of each of the 12 mutations on exit from competence, each mutant reporter strain was treated with CSP at OD$_{550}$= 0.1 under standard conditions and sampled 0, 10, 20, 30, 40, 60, or 80 minutes later for LacZ assay. The enzyme levels, which indicate late gene expression patterns, are shown as a function of the time after CSP induction in Fig. 12. Nearly all of these mutants
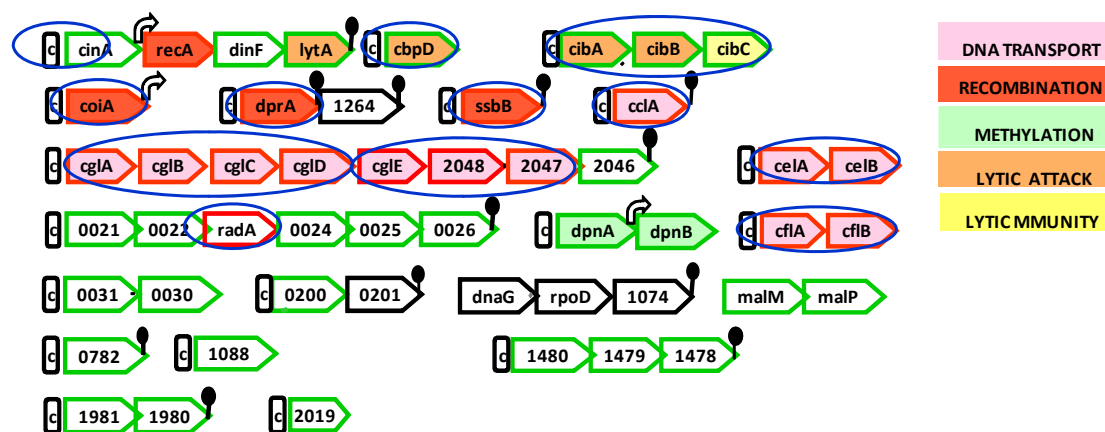
Figure 11 Organization of late genes of the *S. pneumoniae* competence regulon. Internal deletion mutants studied in this paper are indicated by blue ovals. C represents combox. Filled colors symbolize the functions of proteins, as indicated by the key on the right. Bent arrow and lollipop stand for constitutive promoter and terminator respectively. Red border indicate genes essential for transformation. Green border indicate genes not essential for transformation. Black borders indicate genes whose importance for transformation has not been determined. Genes described by numbers (ORF numbers in TIGR4) are those whose functions in competence are unknown. The ORF numbers for these well-defined late genes are: *cinA*(sp1941), *cbpD*(sp2201), *cibA*(sp0125), *coiA*(sp0978), *dprA*(sp1266), *ssbB*(sp1908), *cclA*(sp1808), *cglA*(sp2053), *celA*(sp0054), *cflA*(sp2208). The late gene clusters and associated C-box are as in Peterson et al, 2000, except for cases of apparent read-through transcriptions identified in Oggioni and Morrison (2008).

TABLE IV. COMPARISON OF TRANSFORMATION EFFICIENCY OF NEW STRAINS WITH LITERATURE

| New allele combination | | Relative transformation rate | | |
| | | | Experimental [a] | |
Strain(s)	Mutation	Literature [b]	$ClpP^+$	$ClpP^-$
CP1250, CP2000	Δcps	1		
CPM7	$ssbB^-::lacZ::ssbB^+$	1		
CP1359	$\Delta clpP$::PcTet	1		
CP2108, CP2125	$\Delta comA$::PcErm	1	1	1
CP2109, CP2126	$\Delta cbpD$::PcKan	1	1	1
CP2110, CP2127	$\Delta cibABC$::PcKan	1	1	.4
CP2111, CP2130	$\Delta coiA$::PcKan	.01	.001	.001
CP2112, CP2128	$\Delta cglEFG$::PcKan	0	.001	.001
CP2113, CP2129	$\Delta dprA$::PcKan	0	.0001	.001
CP2116, CP2131	$\Delta celAB$::PcKan	0	.001	.001
CP2114, CP2132	$\Delta cclA$::PcKan	0	.001	.0001
CP2117, CP2133	$\Delta cglABCD$::PcKan	0	.0001	.0001
CP2115, CP2134	$\Delta cflAB$::PcKan	0	.0001	.0001
CP2119, CP2135	$\Delta radA$::PcSpc	0	.001	.001
CP2139, CP2140	$\Delta ssbB$::PcKan	.3	.3	0.5
CP2143, CP2144	Δcombox-$\Delta cinA$::PcKan	1	1	1

(a) Relative transformation rates were calculated by comparing the transformation efficiencies of the mutants to that of CP2000

(b) The literature sources for the relative transformation rates of the mutants are found in Table 1 and/or in Table S2 in Peterson et al. (2004).
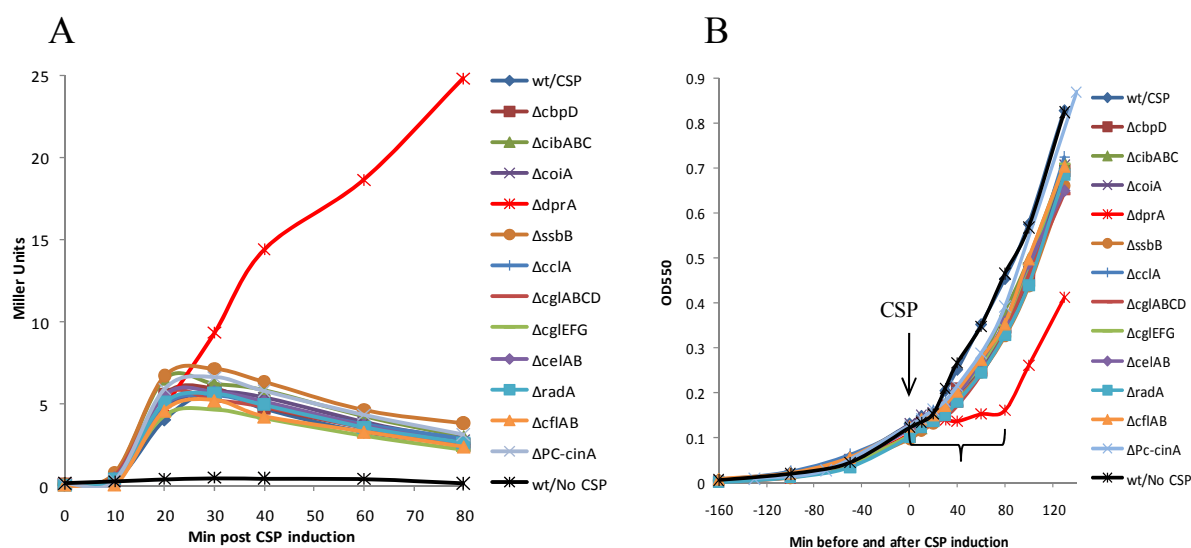
Figure 12 Survey of effect of late gene mutations on exit from the competent state. **A**, Late gene expression patterns were monitored in wild type and late gene mutants after CSP induction. A *lacZ* reporter was inserted at the late gene *ssbB* to show late gene expression. β-galactosidase activity (Miller units) was measured in culture samples taken after CSP induction at indicated times. Each strain is indicated by mutated competence gene (see Table 1). **B,** Growth patterns monitored as culture optical density at 550nm before and after CSP treatment. Bracket indicates period of LacZ assay shown in panel A.

displayed normal (brief) patterns of late gene expression, with a burst of LacZ synthesis restricted to the period between 10 and 30 minutes after addition of CSP. The single exception was the *dprA* mutant, in which late gene expression continued beyond 60 minutes The pattern shown in Fig. 2 reproduces the growth defect reported by Bergé (2002) for the R6 strain, and further supports his interpretation by showing directly that expression of a late gene continues for an unusually long time after exposure of a *dprA* mutant to CSP; that is, the cells appear to remain in an active X-state for a greatly extended period when DprA is missing.

The greatly prolonged expression of a late gene suggests a prolonged presence of ComX and thus that DprA affects not only expression of the *comCDE* operon, but also the *comX* genes. Combined with the report of Mirouze et al. (2007) that premature expression of *dprA* restores the transience defect of a *comX* mutant, these results suggest that DprA is both necessary and sufficient for the shutoff of early gene expression. From the absence of any similar effect of the other late gene mutations on the temporal pattern of *ssbB* expression I conclude further that none of the other late gene products examined has a strong role in shutoff of early gene expression. Because of the uniquely strong effect of DprA on shutoff, it is of interest to determine its effects on gene expression in more detail, both to know if it is the only shutoff agent, and to identify its regulatory targets.

4.3.2 The kinetics of exit from competence is not altered in a ClpP protease-deficient background.

I was concerned to be able to detect regulation of *comX* at multiple levels. While the effect of DprA shown above is dramatic, as expected for a factor controlling early gene expression, the change in late gene expression kinetics might be more subtle for mutation of a gene affecting ComX activity directly. Thus, for all late gene mutants tested above, except the Δ*dprA* mutant, levels of ComX and ComW would be expected to start to decline by ~30 minutes after CSP induction, simply due to the rapid *dprA*-dependent halt to early gene expression and subsequent decay of ComX and ComW that had

accumulated during the brief window of ComE activity. Therefore, any extension of late gene expression occasioned by mutation of a gene acting specifically to suppress ComX activity during this window might be modest in length and difficult to detect. To extend the window of ComX availability and thus improve the chance of detecting such a ComX-dependent shutoff gene targeting ComX itself, I decided to study the temporal pattern of late gene expression in late gene mutants in a protease-deficient background (*ΔclpP*), in which both ComX and ComW would be stabilized. ClpP and ClpE are largely responsible for the proteolysis of ComX and in strains deficient for either of the two proteins, ComX becomes stable (Luo, 2003). Similarly, ClpP and ClpC are largely responsible for the proteolysis of ComW and in strains deficient for one or the other protease subunit, ComW is stable (Piotrowski, et al., 2009).

Adopting the same strategy as in the previous section, a new parent strain, CP2125, was made from the original one by disruption of the *clpP* gene, to increase stability of both ComW and ComX proteins. From this new parent strain, I again obtained 12 late gene mutants: *ΔcbpD, ΔcibABC, ΔssbB, ΔcglEFG, ΔcoiA, ΔdprA, ΔcelAB, ΔcclA, ΔcglABCD, ΔcflAB, ΔradA* and *ΔPc-cinA,* and verified their structures and competence phenotypes as above. These mutants were analyzed for their late gene expression pattern after induction by CSP. In the *ΔclpP* background, I observed the same result as in the protease-proficient wild type: only the *dprA* mutant displayed prolonged late gene expression (Figure 13). Since my survey of ~19 late genes other than *dprA* did not reveal any whose loss extends the X-state when the half-lives of ComX and ComW are prolonged by interruption of their proteolysis, I conclude that none of these gene products is individually responsible for suppression of ComX activity.
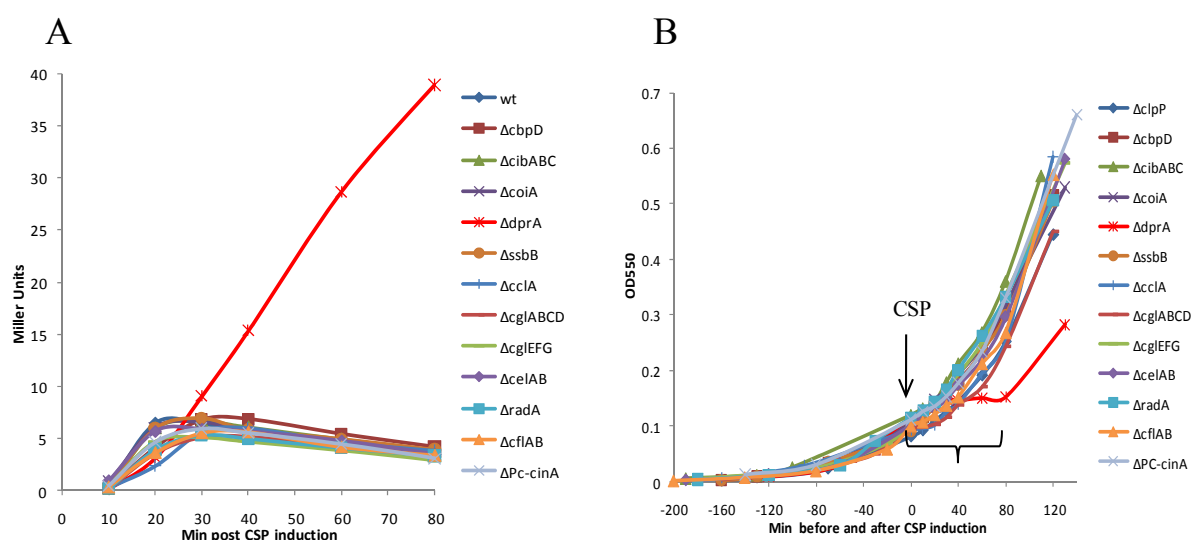
Figure 13. Survey of effect of late gene mutations on exit from the competent state in the protease deficient *ΔclpP* background. A, After CSP induction, late gene expression patterns were monitored in wild type and late gene mutants in *ΔclpP* background. The *lacZ* reporter and the measurement of β-galactosidase activity were done as described for Figure 2. Each strain is indicated by its mutated competence gene as listed in Table 1. **B,** Growth patterns monitored as culture optical density at 550nm before and after CSP treatment. Bracket indicates period of LacZ assay shown in panel A.

4.3.3 CSP induces higher levels of ComX and ComW in a *dprA⁻* background than does raffinose in the ectopic ComX/ComW expression regime.

The prolonged late gene expression in *ΔdprA* mutants might reflect abrogation of a direct inhibitory effect of DprA on late gene expression or on early gene expression. Previous work in ourlaboratory showed that the accumulations of ComX and ComW increased in *ΔdprA* mutants, in both wild type and protease deficient backgrounds (Piotrowski, 2010). This result suggested that the prolonged late gene expression in *ΔdprA* mutants might be due to the continuation of early gene expression which otherwise would be shut off . Moreover, in another *ΔdprA* mutants which contains *comX* and *comW* in the *aga* operon (CP1932, *aga::comX::comW, ΔclpC, ΔclpE, ssbB::lacZ::ssbB⁺, ΔdprA*), competence could be induced by CSP via early gene expression, or by raffinose via ectopic expression of ComX and ComW bypassing early gene expression. While competence induced by CSP was prolonged, competence induced by raffinose displayed normal shutoff. This indicated that DprA does not control late gene expression.

Since late gene expression induced by raffinose was independent of a functional *dprA* gene, I suspect that there may be another regulator, which reverses late gene expression soon after it begins. I further hypothesize that the apparent failure of this inhibition in the *dprA* mutant induced by CSP (Fig. 12 and Fig. 13) could be explained if ComX and ComW were produced in amounts far above those accumulated in wild type competent cells or in the ectopically induced cells, overwhelming the hypothetical regulator. To test this hypothesis, I compared levels of both ComX and ComW in CP1932 induced by CSP to the levels in the same strain induced by raffinose, sampling during 80 minutes after addition of the respective inducers. Two parallel SDS-gels were run, one probed with anti-ComX antibody and another probed with anti-ComW antibody. CSP-induced levels of both ComX and ComW were indeed much higher than those induced by raffinose (Figure 14), even while late gene expression,
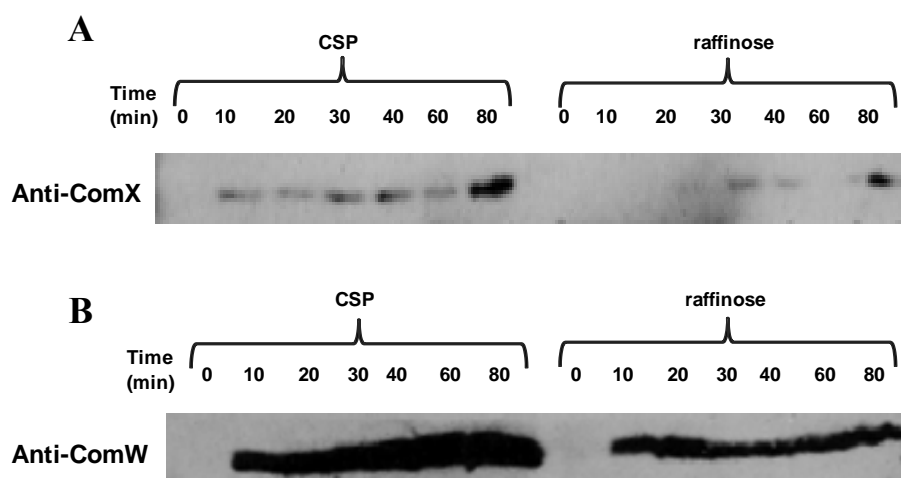
Figure 14 Reduced levels of ComX and ComW under ectopic regulation of competence. Late gene expression was induced in strain CP1932 (*aga::comX::comW*, *ssbB::lacZ*, Δ*clpC*, Δ*clpE*) by 250ng/ml CSP or 0.1% raffinose respectively at OD 0.1. The cultures were sampled at 0, 10, 20, 30, 40, 60, and 80 minutes after induction for western blotting assay. Same amount of samples were loaded into two SDS-PAGE gels running in parallel: one probed with anit-ComX antibody (Panel A), another probed with anti-ComW antibody (Panel B). Similar results were obtained from three repeated experiments.

as represented by a *ssbB* reporter, though prolonged, was not increased in rate (Fig. 12). Thus, the greatly elevated ratio of ComX to a hypothetical late gene product acting as an inhibitor could explain why late gene expression patterns in *dprA* mutants shut off after raffinose induction but not after induction by the native CSP pheromone.

4.3.4 Limited screening of DprA target with yeast two-hybrid assay

Since DprA turns off competence by controlling early gene expression, I started a search of what target it works on. Since ComD or ComE are the major players in turning on early gene expression, I set up yeast two-hybrid assay for a limited screening on these two targets, plus some other controls, such as ComX, ComW, and RecA. Previous work showed that DprA does not affect late gene expression, so *comX* and *comW* inserts were used as the negative controls. We knew that DprA can interact with RecA (Mortier-Barrière et al., 2007), so the *recA* insert was adopted as a positive control. In cloning constructions, I fused *dprA* gene to the DNA binding domain of GAL4 transcription factor in the shuttle plasmid pGBDUC2 and fused *comD*, *comE*, *comX*, *comW* and *recA* to the activating domain of GAL4 in the shuttle plasmid pACT2. The presence of the inserts in plasmids were verified by enzyme digestion (BamHI and BglII), as shown in Figure 15.

After the confirmation of the constructions, pGBDUC2 (BD-ф)and pGBDUC2-*dprA*(BD-A) were transformed into yeast haploid NSY752(α). On the other hand, pACT2 (AD-ф), pACT2-*comD* (AD-D), pACT2-*comE* (AD-E), pACT2-*comX*(AD-X), pACT2-*comW* (AD-W)and pACT2-*recA* (AD-R) were transformed into yeast haploid NSY468(a) (Fig. 16). Yeast Diploids were obtained by mating between NSY752(α)::pGBDUC2-dprA and NSY468(a) cell lines. In total, six diploid cell lines were obtained: BD-A+AD-ф; BD-A+AD-D, BD-A+AD-E, BD-A+AD-X, BD-A+AD-W, BD-A+AD-R. In the diploids, if the activating domain and DNA binding domain could be brought to close proximity by the bait (*dprA*) and prey(*comD*, *comE*, *comX*, *comW* and *recA*), a downstream reporter gene, *HIS3*,
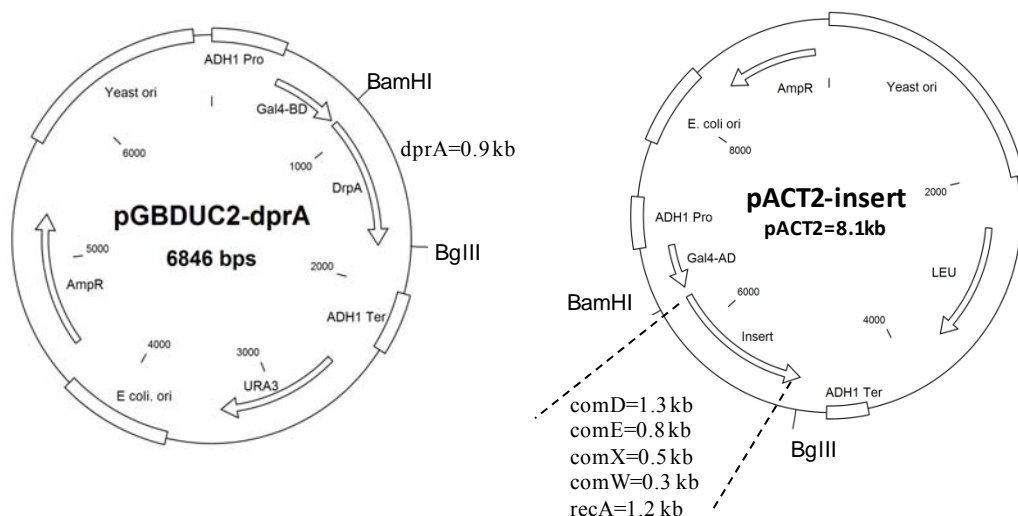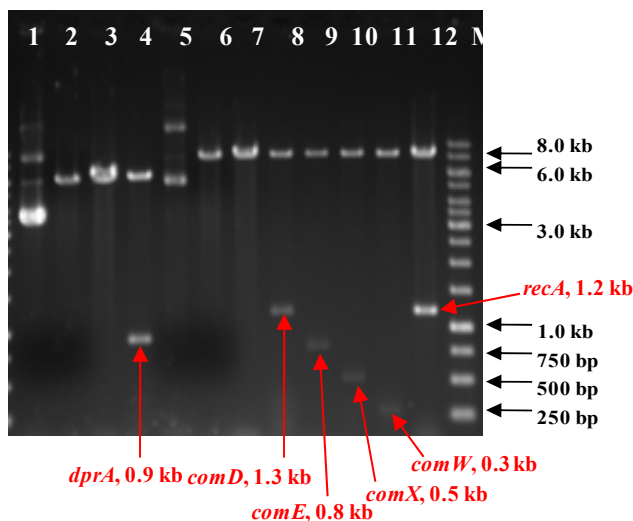
Figure 15 Confirmation of genetic constructions for yeast two-hybrid assay. In genetic construction, *drpA* was inserted into plasmid pGBDUC2 at BamHI and SalI restriction sites. *comD*, *comE*, *comX*, *comW*, *recA* were inserted plasmid pACT2 at BamHI and XhoI sites. To verify their presences in the plasmids, the empty plasmids and hybrid plasmids were digested by BamHI and BglII endonucleases to see the inserts with right sizes could be cut off. About 200 ng plasmid DNA were treated with the enzymes. After the digestions, the reaction mixtures were run on 1% agarose gel for electrophoresis assay. In lane 4, 8, 9, 10, 11 and12. the insert bands were visible with right sizes. This suggested the inserts were in the corresponding hybrid plasmids. pGBDUC2, 6.0 kb; pACT2, 8.1 kb; *dprA*, 0.9 kb; *comD*, 1.3 kb; *comE*, 0.8 kb; *comX*, 0.5 kb; *comW*, 0.3 kb; *recA*, 1.2 kb.
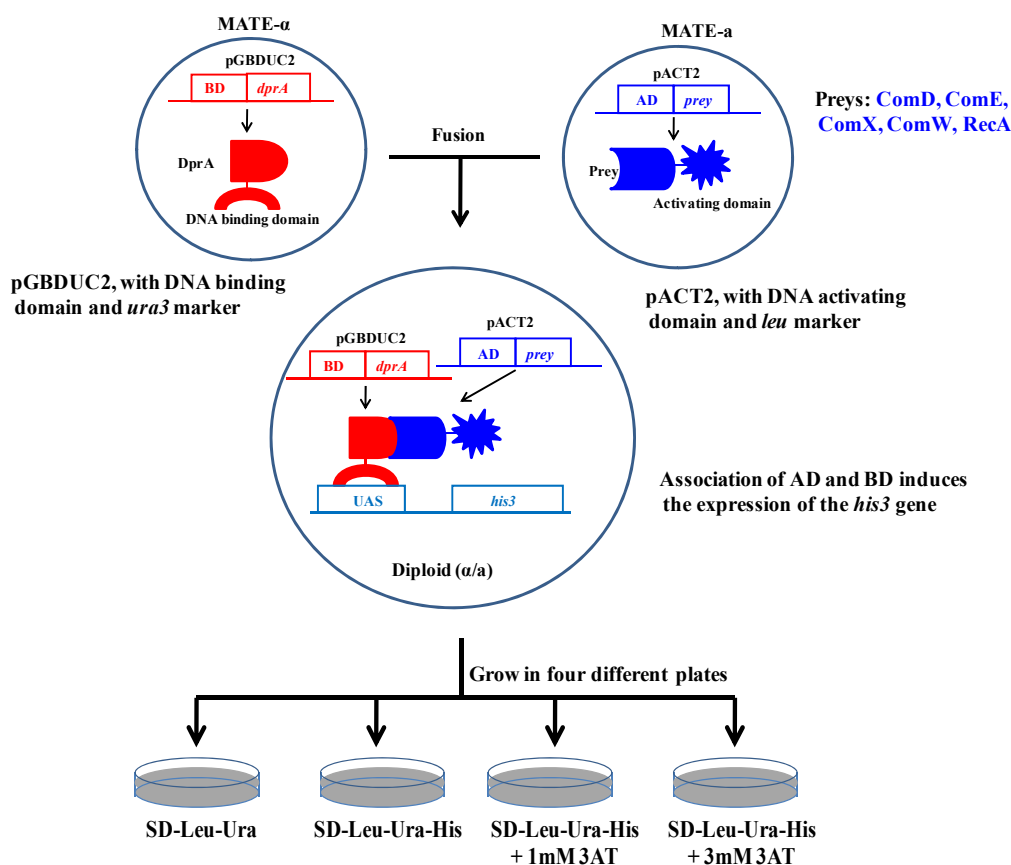
Figure 16 Experimental design of yeast two-hybrid assay for limited screening of DprA's target. *drpA* and preys (*comD*, *comE*, *comX*, *comW*, *recA*) were fused to DNA binding domain and activation domains of Gal4 transcription factor in shuttle plasmids pGBDUC2 and pACT2 individually. Then, the chimeric plasmids were transformed into yeast haploids with opposite mating types. pGBDUC2 (BD-ϕ) and pGBDUC2-*dprA* (BD-A) were transformed into yeast haploid NSY752(α). On the other hand, pACT2 (AD-ϕ), pACT2-*comD* (AD-D), pACT2-*comE* (AD-E), pACT2-*comX* (AD-X), pACT2-*comW* (AD-W) and pACT2-*recA* (AD-R) were transformed into yeast haploid NSY468(a). After cell fusion, diploid cells were collected: BD-A+AD-ϕ; BD-A+AD-D, BD-A+AD-E, BD-A+AD-X, BD-A+AD-W, BD-A+AD-R. Then the diploid cells were grown on four different media. The first plate had selective medium for diploid cells (SD-Ura-Leu). The second plate contained diploid selective medium without histidine (SD-Ura-Leu-His), selecting for the interaction between prey and bait. The third and fourth plates were filled with interaction-selective medium plus increasing amounts of the histidine production inhibitor 3AT (SD-Ura-Leu-His+1mM 3AT and SD-Ura-Leu-His+3mM 3AT), to select stronger interaction by creating more stringent conditions.

would be induced by the recombined GAL4 transcription factor. Therefore, the diploid cells would be able to grow in the medium without histidine.

After the incubation of the diploids in four kinds of media (SD-Ura-Leu, SD-Ura-Leu-His, SD-Ura-Leu-His+1mM 3AT and SD-Ura-Leu-His+3mM 3AT) for certain amount of time, I found all diploids were able to grow in the plate with SD-Ura-Leu (Figure 17). This was what was expected because there are Ura and Leu markers in the plasmids pGBDUC2 and pACT2 respectively. Diploids containing these two plasmids should be able to grow in the SD-Ura-Leu medium. But in the plates without histidine (SD-Ura-Leu-His), only the diploids containing inserts *dprA + comE* and *dprA + recA* were growing up. This indicated that DprA can interact with ComE as much as the way DprA interacts with RecA. The growth in plates without histidine but with the competitive inhibitor of the production of histidine, 3AT, further confirmed the interaction between DprA and ComE is positive and strong, probably stronger than the interaction between DprA and RecA because the diploid with *dprA* and *recA* inserts was unable to grow up in the media SD-Ura-Leu-His+1mM 3AT and SD-Ura-Leu-His+3mM 3AT after a week incubation.

4.3.5 Further verification of the interaction between DprA and ComE with yeast two-hybrid

Since my preliminary screening on DprA targets made ComE stand out, I decided to set up standard yeast two-hybrid assay to verify the interaction between DprA and ComE. Similar strategy was employed as above. Only more negative controls were adopted to rule out the the possibilities that without any inserts, or DprA alone, or ComE alone might bring the GAL4 activation domain and DNA binding domain together. In practice, besides the diploid with both inserts *dprA* and *comE*, I fused the two haploids harboring empty plasmids to make diploid without any inserts (pACT2-ф and pGBDUC2-ф). I also fused the haploids, one harboring empty plasmid and another containing plasmid with the insert *dprA* or *comE*, to make diploids with single insert (pACT2-*dprA* and pGBDUC2-ф; pACT2-ф and

BD-A + AD-D	BD-A + AD-E	BD-A + AD-X	BD-A + AD-W	BD-A + AD-R	BD-A + AD-ɸ

SD-Ura-Leu (Day 2)

BD-A + AD-D	BD-A + AD-E	BD-A + AD-X	BD-A + AD-W	BD-A + AD-R	BD-A + AD-ɸ

SD-Ura-Leu-His (Day 3)

SD-Ura-Leu-His +1mM 3AT (Day 5)

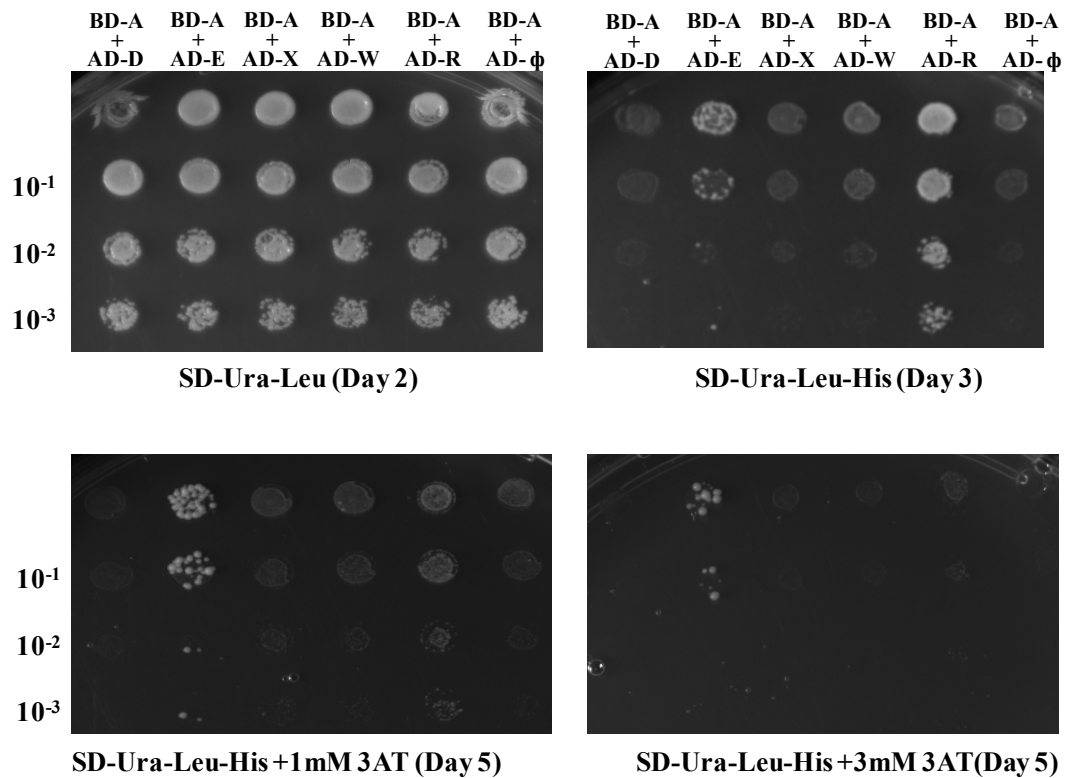**SD-Ura-Leu-His +3mM 3AT(Day 5)**

Figure 17 Limited screening of DprA target with yeast two-hybrid assay. After the incubation of the diploids (BD-A+AD-D, BD-A+AD-E, BD-A+AD-X, BD-A+AD-W, BD-A+AD-R, BD-A+AD-ɸ) in four kinds of media (SD-Ura-Leu, SD-Ura-Leu-His, SD-Ura-Leu-His+1mM 3AT and SD-Ura-Leu-His+3mM 3AT) for certain amount of time, all diploids were found to be able to grow in the plate with SD-Ura-Leu, suggesting they are diploids containing pACT2 and pGBDUC2 plasmids or their derivatives. However, in the plate without histidine (SD-Ura-Leu-His), only the diploids containing inserts *dprA + comE* and *dprA + recA* were growing up, suggesting that DprA can interact with ComE as much as the way as DprA interacts with RecA. In the plates plus 3AT (SD-Ura-Leu-His+1mM 3AT and SD-Ura-Leu-His+3mM 3AT), only the diploid with *dprA* and *comE* inserts was growing well, indicating that the interaction between DprA and ComE is probably stronger that the interaction between DprA and RecA.
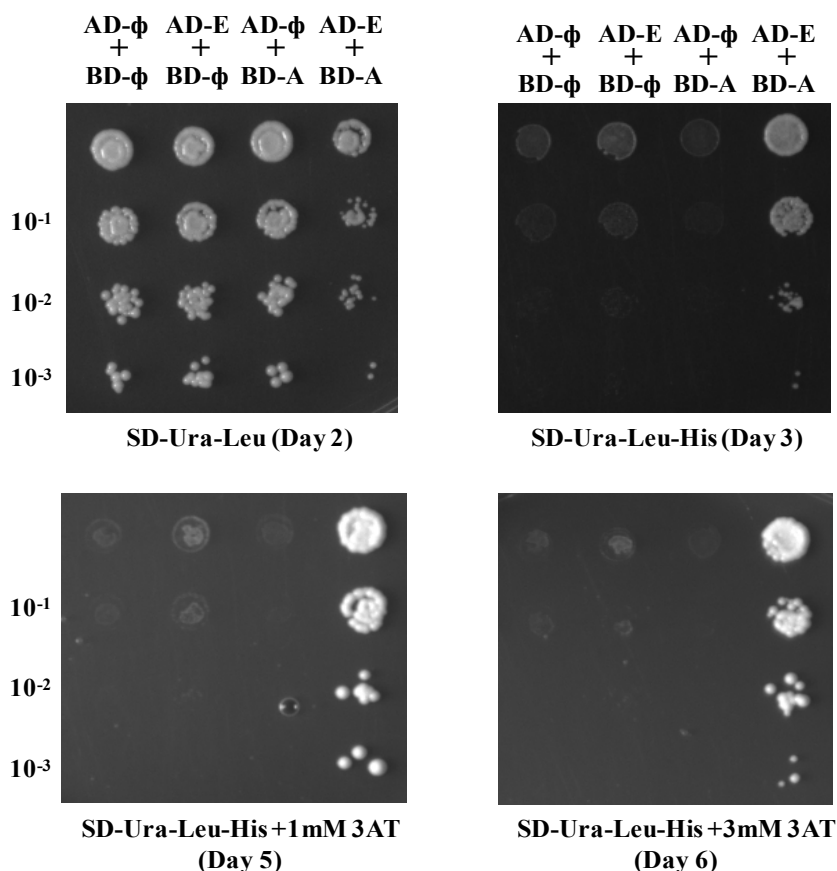
Figure 18 Further verification of the interaction between DprA and ComE by yeast two-hybrid assay. To further confirm the interaction between DprA and ComE, four diploids were obtained by mating NSY468(a) harboring pACT2(AD-φ), pACT2-*comE* (AD-E) and NSY752(α) harboring pGBDUC2 (BD-φ), pGBDUC2-*dprA*(BD-A). After the incubation of the diploids (BD-φ+AD-φ, BD-A+AD-φ, BD-φ+AD-E, BD-A+AD-E) in four kinds of media (SD-Ura-Leu, SD-Ura-Leu-His, SD-Ura-Leu-His+1mM 3AT and SD-Ura-Leu-His+3mM 3AT) for certain amount of time, all diploids were found to be able to grow in the plate with SD-Ura-Leu, suggesting they are diploids containing pACT2 and PGBDUC2 plasmids or their derivatives. However, in the plates without histidine (SD-Ura-Leu-His, SD-Ura-Leu-His+1mM 3AT and SD-Ura-Leu-His+3mM 3AT), only the diploid containing both inserts (*dprA + comE*) was growing well. Diploids containing single insert or no insert (BD-A+AD-φ, BD-φ+AD-E, BD-φ+AD-φ), were unable to grow in media without histidine. This further confirmed that DprA can interact with ComE.

pGBDUC2-*comE*). Eventually, four diploid cell lines were collected: BD-φ+AD-φ; BD-A+AD-φ; BD-φ+AD-E; BD-A+AD-E. Then, these four diploid cells were tested for the growth in same four kinds of plates as above: SD-Ura-Leu, SD-Ura-Leu-His, SD-Ura-Leu-His+1mM 3AT and SD-Ura-Leu-His+3mM 3AT (Figure 18). After incubation at 26ºC for certain amount of time, while everyone was growing normally in SD-Ura-Leu plate, I found only the diploid with both inserts *dprA* and *comE*, not the diploids with single insert or not insert, was able to grow in the SD media without histidine and SD media without histidine but with 3AT. So these results further confirmed the interaction between DprA and ComE.

4.4 Further discussion

The activity of alternative sigma factors is often controlled quite strictly at multiple levels, perhaps because they can cause a global shift in gene expression, which could be especially harmful if carried out under inappropriate circumstances. In the case of the alternative sigma factor central to competence development in *S. pneumoniae*, at least five distinct mechanisms of regulation are already established or glimpsed. (1) Transcription of *comX* depends on a TCSTS that coordinates within local populations and responds to unknown elicitors. (2) Production of ComX protein depends on an early competence gene, *comW*. (3) Separately, activity of ComX depends on ComW. (4) ComX is labile, targeted by the ClpEClpP ATP-dependent protease. (5) ComW is labile, targeted by the ClpCClpP ATP-dependent protease. With the observations reported here on escape from the X state, two additional mechanisms that regulate the activity of ComX can be considered, both apparently forming negative feedback loops that ensure that induction of *comX* expression is self-limiting and transient. One mechanism requires DprA and targets ComE or other determinant of CSP-induced expression of early genes; the other competence regulator is less well defined, but is independent of *dprA* and appears to target ComX or other determinant of transcription of late genes.

Since ComX and ComW are central players during competence development and the disappearance of ComX and ComW occurs at about the same time as the loss of transformation (Peterson, et al 2000; Peterson, et al 2004; Luo and Morrison 2003, and Piotrowski et al., 2009), it is attractive to suggest that the disappearance of ComX and ComW itself accounts for the shutoff of late gene expression (Claverys and Håvarstein, 2002; Claverys et al, 2006). However, the relative timing of events during response to CSP is difficult to reconcile with this simple mechanism. Specifically, late gene mRNA largely disappears even before the levels of ComX and ComW prtoeins begin to drop, suggesting that the activity of SigX is itself subject to some additional form of control (Alloing et al,., 1998; Dagkessamanskaia et al, 2004; Piotrowski et al, 2009; Luo et al., 2003). As some non-proteolytic factor thus seems to play a major additional role in the shut off of ComX activity, I began to look for possible candidates among ComX-induced late genes. Peterson et al. (2004) screened mutations of many of the late genes for effects on the rate of exit from competence, but found none that caused a pronounced extension of the period of transformation. Transformation defective mutants were not examined for this phenotype, however, prompting us to investigate the latter class of late genes directly. *dprA* mutants treated with CSP displayed a prolonged period of expression of late genes, in addition to the prolonged expression of the early gene operon *comCDE* previously described as a mutant phenotype for *dprA* by Bergé (2002) and by Mirouze et al. (2007).

The apparent role of DprA in terminating late gene transcription could thus in principle be either a secondary effect of its inhibition of early gene expression or could reflect an additional direct effect on late gene expression. To see if DprA also had a capacity of turning off late gene expression directly by an effect on ComX, a new strain was created in this study for the ectopic expression of *comX* and *comW*. This new strain can develop competence upon induction of *comX* and *comW* by raffinose treatment. In this strain, the ability to transform was transient despite continued presence of ComX and ComW (Fig.

6), providing a good background to evaluate the possibility of a direct effect of DprA on late gene expression. In a *dprA* mutant derivative of this strain, the pattern of late gene expression following *comX* and *comW* induction perfectly matched the pattern in the *dprA*[+] parent. This strongly suggests that DprA does not suppress ComX activity in late gene transcription, but that another, probably ComX-dependent, gene may be responsible for limiting ComX activity to a short time window.

It remains a challenge to reconcile the different patterns of late gene expression in *dprA* mutants under CSP and raffinose inductions. Specifically, why is late gene expression prolonged in the *dprA* mutant if there is a separate (late) inhibitor directly targeting ComX? I propose that DprA is not the only factor that shuts off competence, but that a second independent repressor accounts for the termination of late gene expression in *dprA* mutants in the ectopic expression system as well as for the prompt termination of late gene transcription in the WT while ComX is still abundant (Luo and Morrison, 2003; Piotrowski et al, 2009). As this second repressor was apparently ineffective in CSP-induced *dprA* cultures, I hypothesize that ComX and ComW are induced at different levels with CSP and raffinose. In the *ΔdprA* mutant induced by CSP, where DprA, the factor normally curbing early gene expression' is removed, there would be a continuous supply of ComX and ComW at elevated levels, so that even if the second repressor could inactivate a normal amount of ComX or ComW, it could be overwhelmed by the unusually high amounts of these regulators. Direct comparison verified this inferred difference in the levels of ComX and ComW achieved under the two expression regimes. The hypothetical factor responsible for the shut off of late gene expression remains unknown. While here I have ruled out ~20 late gene products, several other late genes as well as the entire class of 'delayed' genes remain untested.

Comparison of levels of ComX and ComW in *dprA* mutants vs. wild type and protease defective backgrounds revealed that DprA not only turns off expression of the early gene operon *comCDE*, but

also has a parallel effect on the early genes *comX* and *comW*, thus strongly suggesting an effect on expression of all early genes. The target of DprA action within the regulators of early gene expression is unknown, but interaction with ComD or with ComE or with its DNA or RNA polymerase binding sites would be especially direct paths to permitting exit from the competent state, although it might also reflect a less direct effect of DprA complexes with RecA or with ssDNA. To resolve this puzzle, we set up yeast two-hybrid assay to see if there is any interaction between DprA and its potential target. My results clearly showed that DprA interacted with ComE, not with ComD, ComX and ComW. This finding perfectly explained how DprA controls early genes expression and why competence is a transient process. By the identification of DprA as the competence repressor, a more complete picture of competence regulation in *S. pneumoniae* has been established (Fig. 23).

However, what remains to be established is how DprA inhibits ComE activity via their interaction. Considering that ComE is a response regulator upon phosphorylation, it might be logical to speculate that the association with DprA might block the phosphorylation site(s) on ComE, which makes it unable to be phosphorylated by ComD. What ends up happening is either ComE-DprA complex cannot bind to the direct repeats in front of the promoter of early genes, or still can bind to the direct repeats but cannot recruit the RNAP. More biochemical assays are needed to elaborate this process. To see if the phosphorylation of ComE is affected upon binding to DprA, we could activate the ComDE two-component system by CSP with or without DprA, and compare the structure difference by mass spectrometry. To investigate whether ComE-DprA complex is able to bind to the direct repeat region, DNA footprinting could be run to see if the direct repeat region is still protected in the presence of ComE and DprA. If ComE-DprA could still bind to the direct repeat, in vitro transcription could be conducted to see if the early gene could be transcribed or not in the presence of ComE, DprA, RNAP, and early gene template.

Chapter V. Interaction of ComX and ComW in Development of Competence for Genetic Transformation in *Streptococcus pneumoniae*

5.1 Abstract

Natural competence for genetic transformation in *S. pneumoniae* entails two sets of genes induced by CSP (competence stimulating peptide): the early genes and the late genes. Early gene expression is turned on by ComE, a response regulator that is activated in response to CSP. Subsequently, another early gene product, ComX (the streptococcal alternative sigma factor), switches on expression of late genes, which eventually makes the cell competent. However, ComX does not function alone. A third early product, ComW, is also required for transformation and for late gene expression. However, how ComW participates in competence development is not known. Previous lab work seemed to show that ComW might not be part of RNAP complex, might not interact with ComX and might not be required for the assembly of ComX and core RNAP. Therefore, it was concluded that ComW does not function by interacting with RNAP or ComX but with some other factor in activating ComX. To ask if ComX and ComW interact directly, I took advantage of yeast two-hybrid system by fusing ComX with the activation domain and ComW with the DNA-binding domain of the GAL4 transcription factor. My results revealed a direct interaction between ComX and ComW. Taken together with previous results, my data seem to indicate that the role of ComW in turning on competence might be via a direct interaction with ComX. Considering that ComW is a small protein *S. pneumoniae*, I hypothesize that ComW works as an adaptor for the activation of ComX.

5.2 Introduction

In *S. pneumoniae*, natural transformation requires two sets of gene expression: the early gene expression to produce ComX and ComW and the late expression to make cell competent. Recent research has uncovered several aspects of the regulatory process by which virtually all cells of an unsynchronized pneumococcal culture shift, in concert, from an incompetent state to a fully competent state (Claverys and Håvarstein, 2002). ComX has 159 amino acids and is an alternative sigma subunit of RNA polymerase which is believed to be able to replace the housekeeping sigma factor (σ^A or RpoD) when early genes are induced by CSP (Lee and Morrison, 1999; Luo and Morrison, 2003). ComW was first characterized by using differential fluorescence induction, in which the pneumococcal gene *SP0018* was found be induced by CSP and required for effective transformation (Bartilson et al., 2001). *comW* encodes a small protein with 78 amino acids. Unlike ComX, which has many homologs across species, the sequence of ComW appears to be quite unique. Homology searches revealed that the best match of this protein is the C-terminal of MC085L (791a.a.) in *Mulluscum contagionsum* Virus sutype 1, which shares 31% identity (44% similarity) with ComW and is annotated as a RNA polymerase-associated viral protein (Luo, 2003). Although its in vivo roles remain unknown, evidences have shown that ComW can protect ComX from degradation and activate ComX somehow to induce late gene expression (Sung and Morrison, 2005). When ComX and ComW are ectopically expressed in *aga* operon without turning on early genes, competence can still be induced at similar level of that induced by CSP (Luo and Morrison, 2003; Piotrowski and Morrison, 2010). This suggested that ComX and ComW alone are sufficient to drive the transcription of late genes and they are thus the central players of competence development in *S. pneumoniae*. However, both ComX and ComW transiently exist in the cell during competence development. They are subjected to proteolysis of ClpEP and ClpCP respectively (Luo et

al., 2003; Piotrowski et al., 2010). These features actually add another level of control to competence regulation.

While it is clear that the auto-catalytic peptide pheromone CSP serves to coordinate development of competence among the cells of a culture, the signals or stresses that trigger the developmental cycle are only beginning to be discovered (Claverys et al., 2006; Prudhomme et al., 2000; Martin et al., 2010). Beside ComX and ComW, competence initiation has been found to be affected by some metabolic processes, such as oligo peptide transport, the CiaRH system, purine synthesis and ClpP stress response (Alloing et al., 1998; Claverys et al., 2000; Martin et al., 2000; Claverys and Havarstein, 2002; Robertson et al., 2002). Many external factors have showed to affect competence initiation as well, such as temperature, pH value of the medium, concentrations of metal ions, such as Mg^{2+}, Ca^{2+}, Zn^{2+}, and Mn^{2+} (Hotchkiss, 1954; Tomasz and Hotchkiss, 1964; Chen and Morrison, 1987). It was also found that competence was inhibited if the phosphate source was replaced by maleate (Havarstein and Morrison, 1999). So competence initiation is a delicate process and many factors have to be involved in this process. This partially explains why the competence-specific sigma factor, *comX*, and the ComX-related late genes are widely conserved in many bacterial genomes, but only a few appear to be naturally transformable (Claverys and Martin, 2003; Martin et al., 2006; Mashburn-Warren et al., 2010). In pneumococcus, ComW has to be in place to induce competence.

It was suspected that ComW might be associated with RNA polymerase in *S. pneumoniae*, since the best match of homology search brought up a RNA polymerase-associated viral protein. However, pull-down assays demonstrated that ComW did not co-purify with RNA polymerase even while ComX did (Piotrowski, 2010). An in vitro assembly of RNAP holoenzyme made by expressing and purifying ComX from *E. coli* and core RNAP from *S. pneumoniae* also showed that the assembled holoenzyme is

able to transcribe the late gene templates without the presence of ComW (Luo, 2003). This indicated that ComW might not be absolutely required to load ComX into core RNAP to form the holoenzyme.

It seemed that ComW does not associate with RNAP or participate in the assembly of ComX and RNAP, but it is required for ComX and RNAP activity. So I suspected that ComW might be able to associate with ComX directly to be part of the late gene initiation. To prove this hypothesis, I employed the yeast two-hybrid assay to see if these two proteins would interact in vivo. Strong interaction between ComX and ComW was observed.

5.3 Results

5.3.1 Experimental design of yeast two-hybrid assay

In order to use yeast two-hybrid assay to study interaction between ComX and ComW, two chimeric fusions were created by fusing *comX* and *comW* with the activation domain and DNA binding domain of GAL4 transcription factor respectively. Subsequently, the two chimeric fusions, pACT2-*comX* and pGBDUC2-*comW*, were transformed into yeast haploids with different mating types individually. After cell mating, the yeast diploid cells were plated onto four kinds of media for testing growth: SD-Ura-Leu, SD-Ura-Leu-His, SD-Ura-Leu-His+1mM 3AT and SD-Ura-Leu-His+3mM 3AT. If ComX and ComW can bring the DNA binding domain and the activation domain in close proximity in the yeast diploid, a downstream *HIS3* gene would be transcribed and the diploid would be able to grow in the media without histidine (Figure 19). 3AT, 3-Amino-1,2,4-triazole, is the competitive inhibitor of the production of histidine. The addition of 3AT in the media created more stringent conditions for growing diploids with protein interactions.

5.3.2 Cloning *comX* and *comW* into shuttle plasmids

In practice, *comX* was amplified from genomic DNA of CP2000 with BamHI and XhoI restriction sites, digested, and ligated to the shuttle plasmid pACT2 after its digestion with the BamHI
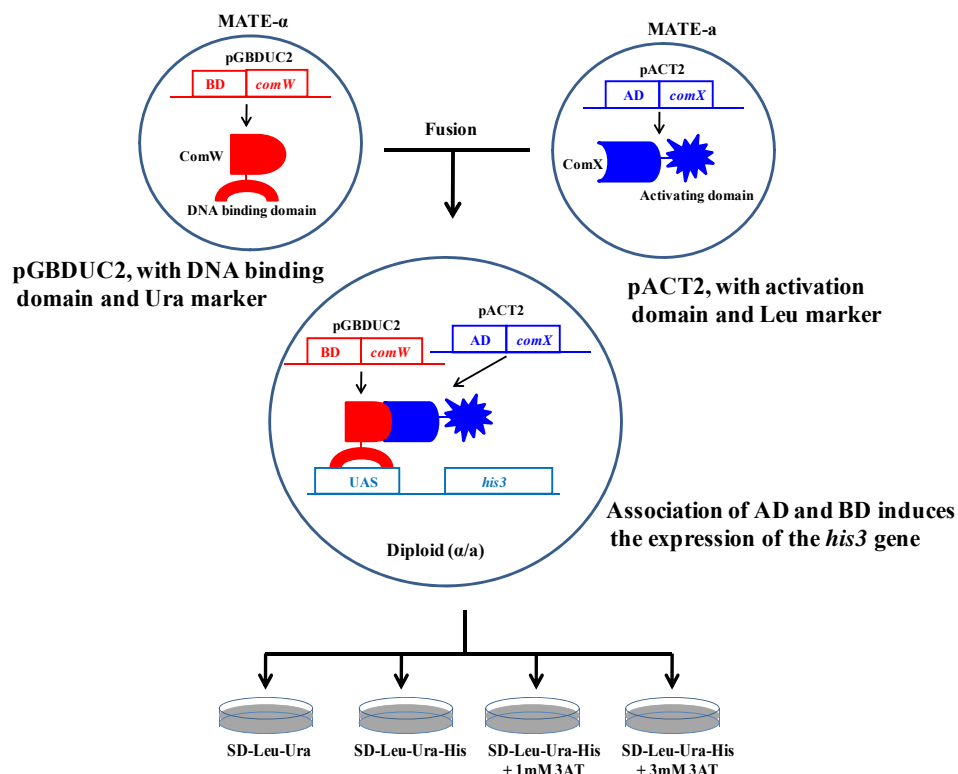
Figure 19 Experimental design of yeast two-hybrid assay for the interaction of ComX and ComW. *comX* and *comW* were fused to DNA binding domain and activation domains of Gal4 transcription factor in shuttle plasmids pACT2 and pGBDUC2 respectively. Then, the chimeric plasmids were transformed into yeast haploids with opposite mating types. pACT2 (AD-ф), pACT2-*comX* (AD-X), pACT2-*comW* (AD-W) and pACT2-*recA* (AD-R) were transformed into yeast haploid NSY468(a). On the other hand, pGBDUC2 (BD-ф) and pGBDUC2-*comW* (BD-W) were transformed into yeast haploid NSY752(α). After cell fusion, four diploid cells were collected: AD-A+BD-ф; AD-ф+BD-W, AD-X+BD-ф, AD-X+BD-W. Then the diploid cells were grown on four different media. The first plate had selective medium for diploid cells (SD-Ura-Leu). The second plate contained diploid selective medium without histidine (SD-Ura-Leu-His), selecting for the interaction between prey and bait. The third and fourth plates were filled with interaction-selective medium plus increasing amounts of the histidine production inhibitor 3AT (SD-Ura-Leu-His+1mM 3AT and SD-Ura-Leu-His+3mM 3AT), to select stronger interaction by creating more stringent conditions.
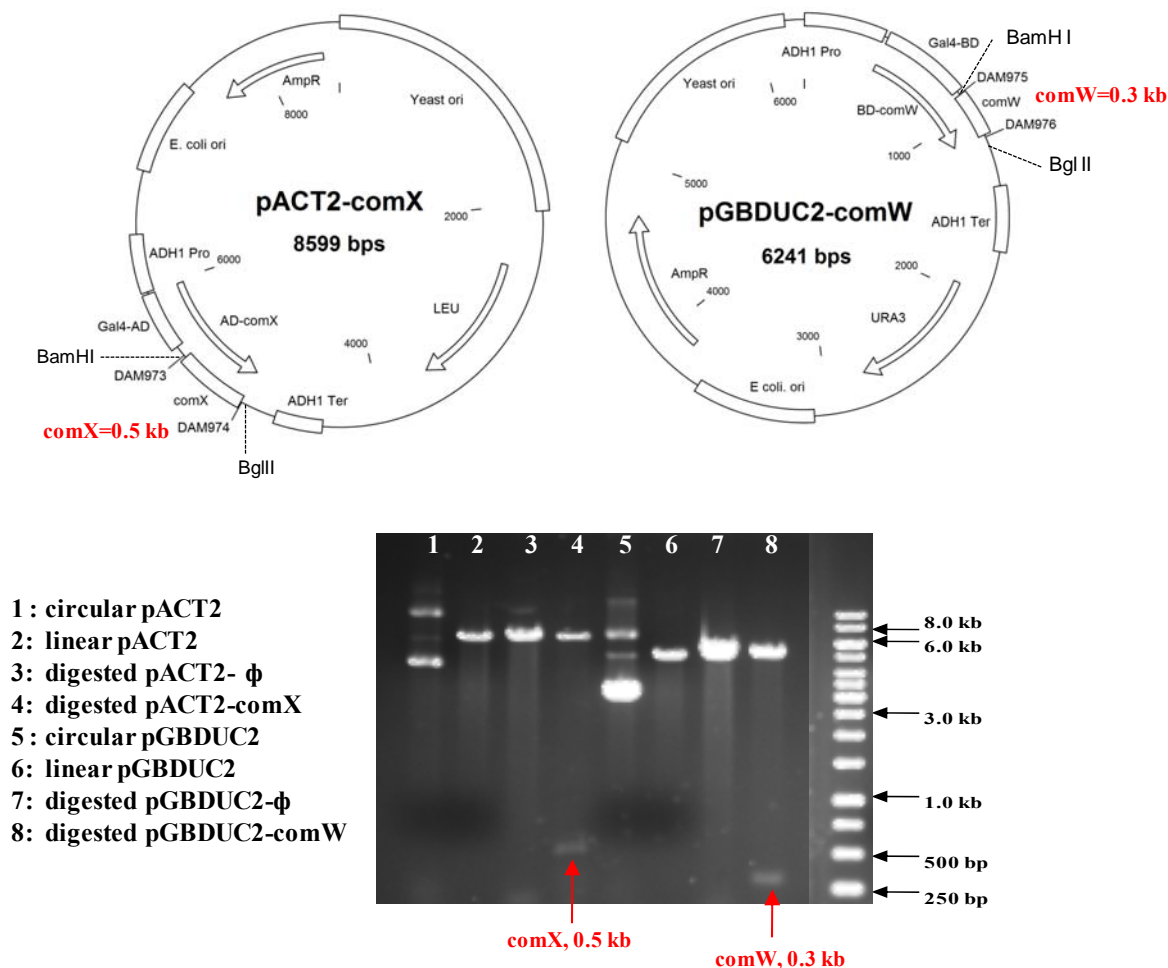
Figure 20 Confirmation of genetic constructions in pACT2-*comX* and pGBDUC2-*comW* for yeast two-hybrid assay. In genetic construction, *comX* were inserted plasmid pACT2 at BamHI and XhoI sites and *comW* was inserted into plasmid pGBDUC2 at BamHI and SalI restriction sites. To verify their existences in the plasmids, the empty plasmids and hybrid plasmids were digested by BamHI and BglII endonucleases to see if the inserts with right sizes could be cut off. About 200 ng plasmid DNA were treated with the enzymes. After the digestions, the reaction mixtures were run on 1% agarose gel for electrophoresis assay. In lane 4 and 8, two bands were visible in lower weight: 0.5 kb (*comX*) and 0.3 kb (*comW*), which indicated the presence of *comX* in pACT2 and *comW* in pGBDUC2. pGBDUC2, 6.0 kb; pACT2, 8.1 kb; comX, 0.5 kb; comW, 0.3 kb.

and XhoI endonucleases. On the other hand, *comW* was amplified from genomic DNA of CP2000 with BamHI and SalI restriction sites, digested, and ligated to the shuttle plasmid pGBDUC2 after the digestion with BamHI and SalI. The two chimeric plamids, pACT2-*comX* and pGBDUC2-*comW*, were subsequently transformed into *E. coli* cell line (DH5α) for enrichment. The presence of *comX* in pACT2-*comX* and the presence of *comW* in pGBDUC2-*comW* were verified by digestion of the hybrid plasmids with BamHI and BglII. After the digestion, the cutoff insert bands were visible on the gel, a 0.5-kb band in lane 4 and a 0.3-kb band in lane 8, which are corresponding to *comX* and *comW* respectively. This result proved the presence of *comX* in pACT2 plasmid and *comW* in pGBDUC2 plasmid (Figure 20).

5.3.3 Transforming shuttle plasmids into yeast haploids

The yeast haploid cell line, NSY752(α), was transformed individually with plamid DNA of pACT2 and pACT2-*comX*, and selected on the synthetic medium without leucine (SD-Leu). Meanwhile, another yeast haploid cell line, NSY468(a), was transformed with plasmid DNA of pGBDUC2 and pGBDUC2-comW, and selected on the synthetic medium without uracil (SD-Ura). Colonies were picked from the selective plates after the incubation at 26ºC for 3 days. Eventually, four kinds of yeast haploids were collected: NSY752(α)::pACT2 (AD-ɸ); NSY752(α)::pACT2-*comX* (AD-X); NSY468(a)::pGBDUC2 (BD-ɸ); pGBDUC2-*comW* (BD-W).

5.3.4 ComX interacts with ComW in yeast 2-hybrid assay

By cross-mating between different mating types of the four haploids, four yeast diploid strains were obtained: AD-ɸ + BD-ɸ, AD-ɸ + BD-W, AD-X + BD-ɸ, AD-X + BD-W. Then the diploid cells were grown on four kinds of plates with different media. The first plate (SD-Ura-Leu) has selective medium for diploid cells, where all diploids are supposed to grow. The second plate was filled with diploid selective medium without histidine (SD-Ura-Leu-His). Therefore only the diploid cell having the interaction between prey and bait can grow on it. The third and fourth plates contain interaction-
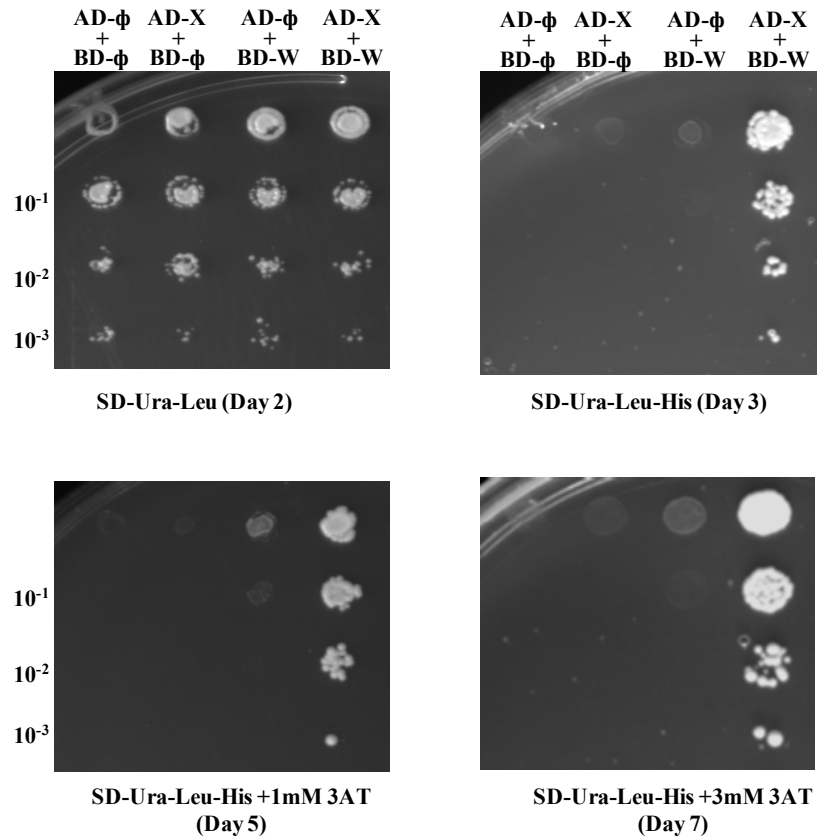
Figure 21 Exploration of the interaction between ComX and ComW by yeast two-hybrid assay. To investigate the interaction between ComX and ComW, four diploids (AD-A+BD-ф; AD-ф+BD-W, AD-X+BD-ф, AD-X+BD-W) were incubated in four kinds of media (SD-Ura-Leu, SD-Ura-Leu-His, SD-Ura-Leu-His+1mM 3AT and SD-Ura-Leu-His+3mM 3AT) for certain amount of time. While all diploids were found to be able to grow in the plate with SD-Ura-Leu, only the diploid containing both inserts (*comX*+*comW*) was growing in plates without histidine (SD-Ura-Leu-His, SD-Ura-Leu-His+1mM 3AT and SD-Ura-Leu-His+3mM 3AT). Diploids containing single insert or no insert (AD-X+BD-ф, AD-ф+AD-W, AD-ф+BD-ф), were found unable to grow in media without histidine. This suggested that ComX can interact with ComW.

selective medium plus different amount of the histidine production inhibitor, 3AT. So, they are used to select strong interaction by creating more stringent conditions.

All the diploids were found to be able to grow in the first plate (Fig. 21), indicating that they were the diploids as expected. But only the diploid with both *comX* and *comW* inserts was able to grow in the second, third and fourth plates with media without histidine, or without histidine but plus 3AT. This strongly suggested the interaction between ComX and ComW.

5.6 Further discussion

The expression of ComX-dependent late genes is necessary for competence development in bacteria. However, the co-existence of *comX* and late genes in the genome cannot guarantee the ComX-dependent late gene expression and competence development, as it has been shown that homologs of *comX* and lates genes are present in many *Streptpcoccus* groups but only few of them displayed natural transformation. To make late gene expressed, the alternative sigma factor ComX has to be successfully recruited into core RNP polymerase. This process may not go as smoothly as expected. It is really logic to speculate that the substitute of ComX for the primary sigma factor (σ^A) might entail some adaptor or activator to be in place. In *S. pneumoniae*, ComW has been found to be required for efficient transformation, protecting ComX from degradation and activating ComX to be functional. It is very likely that it might fit into this role.

However, there is nothing really known about ComW at this point. It is a very unique protein. By sequence search, the best match indicated that it shares low similarity with the C-terminus of a RNA polymerase-associated viral protein. So its possible association with RNAP was investigated by pull-down assay. Its inability to co-purify RNAP indicated that ComW might not associate with RNAP. At first glance, this finding seemed to be at odds with my data which showed the interaction between ComX and ComW, and the canonical theory that ComX could interact with RNAP. But on second

thought, this seeming contradiction could be resolved by two possibilities: 1) the observation that ComW did not co-purify with RNAP does not mean there is absolutely no association between ComW and RNAP, since weak or transient interaction usually cannot be detected by this way; 2) conformational change occurred to ComX when it gets recruited into RNAP, which makes it unable to interact with ComW anymore.

Since the link between ComX and ComW has been set up, it seems to us that ComW is able to modify ComX. So I wonder if there is any post-translational modification made by ComW during the activation of ComX (Hypothesis I in Figure 22). If so, this modification could be phosporylation, or methylation, or acetylation, or others. To explore these possibilities, I purified ComX in both *comW* proficient and deficient backgrounds and submitted them for mass spectrometry assay to see if there is any structural difference. If there is a difference, the mass spectrum might be able to show us what and where the modifications are. Eventually, this experiment would give us a clue what happens between ComX and ComW during their interaction.

Could it be possible that ComW facilitates the assembly of ComX and core RNAP in late gene expression? To test this hypothesis, in vitro transcription was conducted by previous lab mate (P. Luo, 2003), with ComX purified from *E. coli* and core RNAP purified from *S. pneumoniae*. It was found that the in vitro assembled RNAP-ComX complex was able to transcribe late genes without assistance of ComW. This suggested that ComW might not be required to load ComX into core RNAP. However, there are at least two possibilities to account for this. First, even though there is no ComW in *E. coli*, that does not mean there is no protein performing similar function as ComW in *E. coli*. It is possible that ComX got activated already when it was purified in *E. coli* by some unknown protein. Second, the situations were different in vivo and in vitro transcriptions. In vitro, the situation was simplified: there are only ComX and core RNAP in the niche. But in vivo, there are many other interactive proteins and

sigma factors floating aour025nd. For example, the house-keeping sigma factor σ^A is always expressed. It might occupy the core RNAP, which makes ComX hard to have access to core RNAP. The procedure of switching gears for RNAP might need help of some adaptors. So, it is possible that ComW might play a role in facilitating the falloff of σ^A from the core RNAP, therefore contributing to the in vivo transcription but not to in vitro transcription (Hypothesis II in Figure 22).

To discriminate between these two hypotheses, two experiments could be conducted. First, we could purify ComX from both *comW* wild type and mutant backgrounds and look to see if there is any structure difference using mass spectrometry to detect post-translational modifications. If there is a difference, it would be consistent with Hypothesis I. Then, a close examination of the mass spectrum might shed light on what kind of modification was made and on what residue(s). If no difference was found, Hypothesis II might make more sense. To investigate that, in vitro transcription could be employed to study the role of ComW in affecting the competition between ComX and σ^A in late gene transcription. However, Hypothesis I and II do not have to be exclusive to each other. They could be unified if ComW could modify ComX somehow and help it to take the position of σ^A. In this scenario, Hypothesis I and II could be both right.
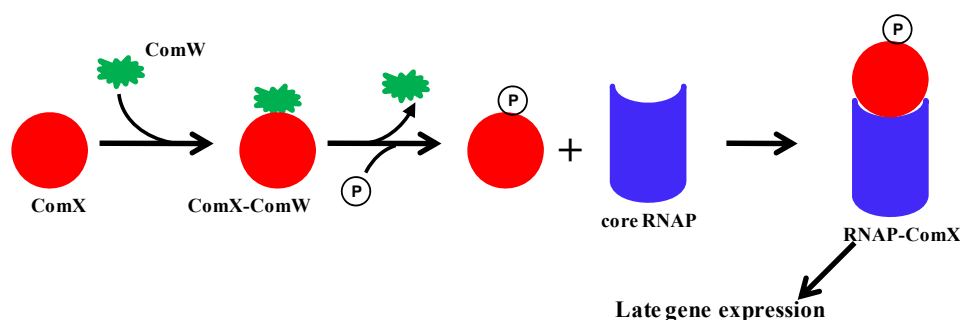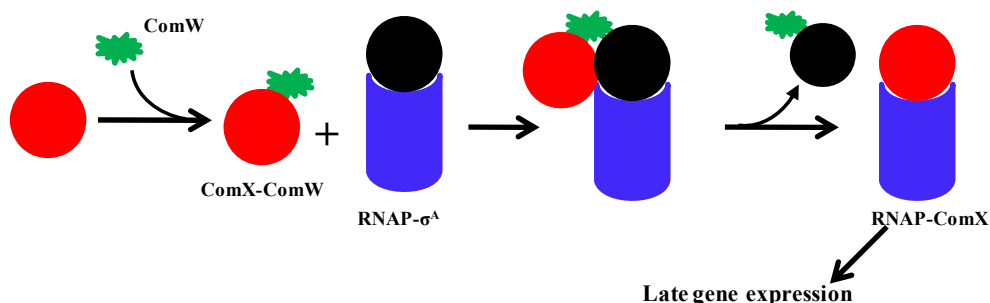
94

Figure 22 Working models for the association between ComX and ComW in turning on late gene expression. ComW is required for the RNAP-ComX activity in vivo. One possibility is that ComW might modify ComX somehow to activate it be functional as a sigma factor (Hypothesis I). Another possibility is ComW might help ComX to take the position of σ^A in RNAP, therefore help ComX to outcompete σ^A in the assembly of RNAP complex (Hypothesis II). P in circle indicates some kind of post-translational modification being made to ComX, such as phosphorylation, acetylation, or acetylation.

Chapter VI. General Conclusion and Discussion

At the beginning of this Ph.D. program, I proposed three specific aims in the preliminary exam: 1) set up the markerless multiple gene deletion system in the application of *Streptococcus pneumoniae*; 2) make multiple deletions of the unknown but highly conserved late genes to investigate their functions; 3) search for the putative ComX-dependent repressors on the shut-off of competence. The first and third aims had been successfully achieved. The second aim was suggested to focus on the DprA story. After following the suggestions from the advisory committee, I found DprA's target in controlling early expression. And I also obtained a bonus finding about the interaction between ComX and ComW which clarified many confusions related to competence initiation in *S. pneumoniae*. Endeavors have also been made to elucidate how ComX is activated by ComW.

As a naturally transformable species, pneumococcus is especially accessible to genetic manipulation. To study its competence regulation, it is sometimes desirable to construct strains with double, triple, or multiple gene disruptions. This is especially significant when it comes to exploration of those uncharacterized late genes. Many of them are highly conserved but their functions in competence development are completely unknown. Even though single disruptions had been made on each of them, none displayed affected competence. A long- unanswered question is if they are required for competence at all, or if there is any redundant functions among them. To answer these questions, multiple disruptions of these genes are required. However, when traditional approach was employed to make multiple gene disruptions in genetic studies, experimental design was often restrained by limited numbers of antibiotic-resistant markers, as well as by the worry of causing accumulation of heterologous proteins in the resulting mutant. So my first objective in this program was to set up a markerless gene deletion system in *S. pneumoniae,* detailed in Chapter III. A *lox66-erm*-t1t2-P$_{fcsk}$-*cre-lox71* cassette,

named Cheshire, was constructed by taking *lox* sites and *cre* gene from Bacterialphage P1, taking *erm* from *Streptococcus*/*E.coli* plasmid pVA838, taking t1t2 from *E. coli* rRNA genes, and taking P_{fcsk} from the fucose operon of pneumococcus. This cassette can be subsequently fused to the upstream and downstream sequences of the target gene and transformed into pneumococcal cells as donor DNA. Once integrated into the genome, the recombinase gene, *cre*, could be induced with the treatment by fucose. With the accumulation of Cre recombinase in the cytoplasm, the crossover between *lox66* and *lox71* would be catalyzed and the internal sequence of the two *lox* sites would be popped out, including the resistance selective marker, *erm*. Eventually, the target sequence disappeared, with only a 34-bp lox72 site left behind, as reminiscent of the Cheshire cat in *Alice's Adventures in Wonderland* (Lewis Carroll, 1896).

While a variety of marker-less gene disruption strategies have been developed for application in pneumococcus，such as the SOEing PCR plus PCR screening (Iannelli *et al*. 2004 and Standish *et al*. (2005), the drop-in/pop-out of disruptive plasmid (Kloosterman et al, 2006), and the JANUS cassette (Sung et al., 2001), they involve either multiple genetic manipulation steps or tedious screening steps. The Cre/*lox* system has enabled efficient routes to making marker-less deletions for broad applications in a large variety of species, including plants, the mouse, yeasts and bacteria. Therefore, introducing Cre/loxP system into pneumococcus for genetic manipulation had become increasingly appealing and theoretically plausible. However, current bacterial implementations of Cre/loxP system typically employ a four-step strategy of (I) emplacing the selectable marker and single-mutant *loxP* sites as a substitute for the intended deletion target, (II) introducing a heterologous *cre* gene into the resulting mutant, (III) allowing expression of *cre* and excision of the selectable marker, and (IV) removing the heterologous *cre* gene from the resulting deletion mutant. To simplify this process for use in pneumococcus while taking advantage of its highly efficient natural transformation system, we sought to combine steps (I)

and (II) and obviate step (IV) by creating a new self-deleting *lox66/erm/cre/lox71* cassette (Fig. 10). Because the excision would be irreversible, the level of expression of the *cre* gene in such a cassette should in principle be low enough to maintain the stability of Cheshire cassette in the genome for steps (I) and (II), but high enough to allow genomic instability to pop out Cheshre cassette at step (III) and (IV). A native regulated pneumococcal promoter (P_{fcsk}), which has a low basal level of expression but is readily activated when is treated with fucose, was chosen to control the expression of *cre* gene and proved to be reliable. To broaden its application in other bacteria, we can just simply replace the fucose promoter with other inducible promoters, or include *fcsR* in the cassette or other place in the genome to use this *lox/erm/cre/lox* design. So far, more than 30 laboratories around the world have requested the Cheshire cassette from us and at least 2 successful applications have been published at this point (Fontaine et al., 2010; Tran et al., 2011).

Both from the angles of transcription and translation, it has been shown that competence development in *S. pneumoniae* is a tightly controlled process, suddenly appearing and rapidly disappearing. This sudden appearance is initiated by an alternative sigma factor, ComX, which can induce a serial of late gene expression simultaneously. However, ComX cannot fulfill this mission alone. Another protein ComW, which is co-expressed with ComX, also plays a critical role in late gene expression. Although not well defined, ComW has been found to protect ComX from degradation and activate the ComX-RNAP complex to be functional. Previous lab work had tried to investigate the relationship between ComW and core RNAP. It was observed that ComW did not co-purify with core RNAP as the same way as ComX does. In vitro transcription assay further showed that ComW might not be required to load ComX into core RNAP to transcribe late genes. These data seemed to tell us that ComW might not interact with core RNAP directly. Usually, refuting one possibility raises another one. The association between ComX and ComW had been long suspected but not established. So I tried to

answer this question with yeast two-hybrid assay and it turned out be a success. Since then, a more detailed picture had been established: ComW interacts with ComX to make it activated and ComX interacts with core RNAP to turn on late gene expression (Figure 23, green arrow).

However, one cannot conclude that the interaction between ComX and ComW is strong, because the method I used was the conventional yeast two-hybrid assay. It did not show quantitative information. What's more, there was no positive control to make comparison. One way to improve this dataset could be adopting the quantitative yeast two-hybrid assay, in which a *lacZ* gene is employed as the substitute of *HIS3* to report protein interaction. The subsequent β-galactosidase liquid assay would tell us how strong the interaction is (Kamiya et al., 2010).

Actually, binding affinity is not the only factor affecting the result of yeast two-hybrid screens. Protein abundance also plays a role in this assay. Method-specific biases have been observed in studying protein interactions. It was reported that protein abundance is the most important factor for detecting interactions in tandem affinity purifications, while it is of less importance for yeast two-hybrid screens (Bjorklund et al., 2008). So, in most cases, protein interactions observed in yeast two-hybrid assays reflect their binding affinities. But the best way to show binding affinity is the surface plasmon resonance.

To further explore the interaction between ComX and ComW in vitro, as well as to find out their binding affinity, I was trying to use SPR to investigate their association and dissociation on the sensor-chip surface. Due to the insolubility of ComW during purification, I tried using MBP-ComW as a substitute to study the interaction with ComX, while using MBP alone and ComW as the control. Unfortunately, I did not see any binding behavior from the sensogram. As always, negative results could hardly confirm or refute anything. No binding affinity between ComX and MBP-ComW does not mean ComX and ComW cannot asociate in vitro. Considering MBP is five times bigger than ComW, it is

quite possible that the interacting face of ComW and ComX might be blocked by MBP. Furthermore, I just tried the CM5 sensor chip. This chip may not be a good choice for this assay. SA chip (for biotinylated protein) or NTA chip (for His-tag protein) might have the potential of producing better results in the future endeavors.

Another unanswered question is how ComX is activated by ComW. In Chapter V, I proposed two hypotheses: one is that ComX might be modified and activated by ComW; another possibility is ComW might work as an adaptor helping ComX outcompete σ^A in loading onto RNAP. More biochemical assays are needed to have a clear answer to this question, such as the analysis of post-translational modification of ComX by mass spectrometry and in vitro transcription.

As a capability of taking up foreign DNA into its own genome, competence has helped diversify this species, gain new functional genes and adapt changing environment as it evolves. However, as a stressful response involved in the production of bacteriocin and hydrolase, competence development might bring extra pressure to the whole population and be especially harmful to the non-competent cells. So the best way to balance these concerns is to let the competence transiently exist: maintaining the capacity of taking up new properties while limiting adverse effect to the whole population as much as possible. In *S. pneumoniae*, natural transformation lasts about 30 min and rapidly shut off. During this period, no cell lysis was observed in wild type strains, which implied that this temporary timing has been well evolved.

Natural transformation in *S. pneumoniae* entails two sets of gene expression: the early genes and late genes. ComX and ComW are the link between these two sets of gene expressions and have proved to be the central players during competence development. Therefore, the termination of competence could be from three possible actions: turning off early gene expression, affecting ComX and ComW stability, and turning off late gene expression. First, ComX and ComW are early gene products and

dependent on early gene expression. Their presence in the cell is subject to CSP-coordinated quorum sensing mechanism and ComDE-mediated signaling transduction. So, turning off early expression is like cutting off ComX and ComW from the source. Second, both ComX and ComW are labile, targeted by proteolysis machinery ClpEP and ClpCP respectively. Their protein levels correspond with competence development. These protease machines impose extra level of competence control. Third, as a sigma factor, ComX needs to get recruited into RNA polymerase to turn on late gene expression. Its function depends on ComW activity. Late gene expression could be repressed by affecting the assembly of ComX and core RNAP, or by disrupting ComW activity. But which one is the primary force shutting off competence? Our lab had suspected that the instability of ComX and ComW might be the primary cause of competence termination. But later evidence showed that this hypothesis might not be the case due to the fact that competence still shut off when ComX and ComW were stabilized. So it is more likely that the instability of ComX and ComW serves as the backup plan: when other competence terminators failed, the protease machinery would kick in.

The shutoff of competence was also observed in the strain with the ectopically expressed *comX* and *comW*, when early gene expression was bypassed. Evidence had pointed out that there might be a late gene product that shuts off competence primarily. This became one of the objectives in my studies, to look for the ComX-dependent repressor from late genes. The transformation-essential late genes became the subject of my studies due to two thoughts: 1) the transformation-nonessential late genes had been thoroughly explored by previous lab work and none of them seemed to be able to shut off competence; 2) turning off competence in time is a critical step in the cell, therefore this action is more likely be conducted by a candidate from transformation essential genes. However, the challenge of studying competence in this category of late genes is that there is no competence development if any of them is disrupted. To circumvent this cumbrance, a *lacZ* reporter was placed in a late gene operon, *ssbB*,

to indicate late gene expression, since late gene expression has been a reliable indicator of competence development. After screening more than 20 late genes in 11 operons, we were able to show that only *dprA* is required for normal shutoff of late gene expression in both wild-type and protease deficient backgrounds. Further characterizations of DrpA targets on competence termination indicated that early gene expression is under the control of *drpA* expression. This finding provided a reasonable explanation on the rapid disappearance of competence and completed the mechanism of competence development in *S. pneumoniae*. Actually, a repressor from late gene expression shutting off early gene expression makes more senses than other scenarios. If the repressor was from early gene expression or proteolysis machinery, whether it shut off early or late gene expression, competence could hardly be developed because of the immediate presence of inhibitor. Taken together, I proposed that DprA is the primary force of competence termination.

Since DprA had stood out as the ComX-dependent repressor turning off early gene expression, we went further to ask what target it works on. Through yeast two-hybrid assay, we observed the interaction between DprA and ComE, but not ComD. This suggested that DprA might inhibit early gene expression by interacting with ComE (Figure 23, green lines). ComE has appeared to be a response regulator upon the phosphorylation from ComD. It is very likely that the binding of DprA to ComE makes it unable to be phosphorylated and, therefore, unable to turn on early gene expression, either by making it unable to bind to the direct repeat in front of early gene promoter, or by blocking the recruitment of RNAP to promoter region. This hypothesis needs to be verified in future experiments, which are detailed in the discussion of Chapter IV.

Is there any factor controlling the expression of late genes directly? I think answer is YES. In an effort to figure out if DprA also inhibit late gene expression, a new strain was created by previous student which decouped early gene expression with late gene expression by ectopically expressing *comX*

and *comW*. This new strain can develop competence upon induction of *comX* and *comW* by raffinose treatment. In a *dprA* mutant derivative of this strain, the pattern of late gene expression following *comX* and *comW* induction perfectly matched the pattern in the *dprA*$^+$ parent. This suggested that DprA does not suppress ComX activity in late gene transcription, but that another, probably ComX-dependent, gene may be responsible for limiting ComX activity to a short time window.

The hypothetical factor responsible for the shut off of late gene expression remains unknown. So far, the late genes have been thoroughly tested on competence termination, including transformation-essential and transformation-nonessential, only DprA stood out controlling early gene expression. It is more likely that the factor regulating late gene expression might come from other categories of competence genes, such as the delayed genes and repressed genes. A genetic screening on these genes would probably give us a surprise.
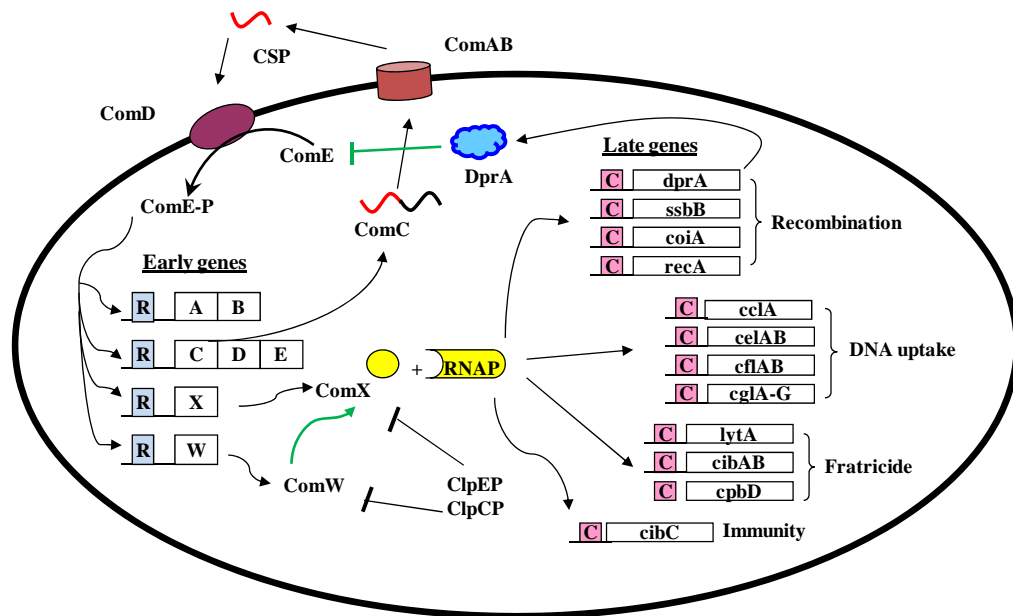
Figure 23 The perspective of competence regulation in *S. pneumoniae*. Competence development in *S. pneumoniae* entails the expressions of two sets of genes, the early gene and late genes. The early gene expression is auto-catalyzed by a positive feedback loop composed of *comAB* and *comCDE*. The overall effect of early gene expression is to secrete CSP to the environment and accumulate ComX and ComW in the cytoplasm. Via direct association between ComX and ComW, ComX becomes able to get recruited into RNAP and turn on late gene expression, which finally make cell competent. One of the late gene product, DprA, beside its function in DNA recombination, can interact with ComE and turn off early gene expression almost immediately after competence was developed. This is primary force shutting off competence. Competence development is also controlled by the proteolysis machinery, ClpEP and ClpCP, like a scavenger who comes into play to clear out ComX and ComW ultimately when competence is terminated by DprA. All these come together to make sure competence in *S. pneumoniae* is a transient process and tightly controlled. Green lines indicate my contributions to the competence story in pneumococcus.

CITED LITERATURE

Albert, H., E. C. Dale, et al. 1995. Site-specific integration of DNA into wild-type and mutant *lox* sites placed in the plant genome. Plant J. 7:649-59

Alloing, G., B. Martin, et al. 1998. Development of competence in *Streptococcus pneumonaie*: pheromone autoinduction and control of quorum sensing by the oligopeptide permease. Mol. Microbiol. 29:75-83

Antoine, D. S., G. Christophe, et al. 2000. Microarray-based identification of a novel *Streptococcus pneumoniae* regulon controlled by an autoinduced peptide, J. Bacteriol. 182:4696–703

Ansaldi, M, D. Marolt, et al. 2002. Specific activation of the Bacillus quorum-sensing systems by isoprenylated pheromone variants. Mol. Microbiol. 44:1561-73

Attaiech, L., C. Granadel, J. P. Claverys, and B. Martin. 2008. RadC, a misleading name? J. Bacteriol. 190:5729-32

Avery, O. T., C. M. MacLeod, and M. McCarty. 1944. Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. J. Exp. Med. 79:137-58

Banerjee, A., and I. Biswas. 2008. Markerless multiple-gene-deletion system for *Streptococcus mutans*. Appl. Environ. Microbiol. 74:2037-42

Bartilson, M., A. Marra, et al. 2001. Differential fluorescence induction reveals *Streptococcus pneumoniae* loci regulated by competence stimulatory peptide. Mol. Microbiol. 39:126-35

Bassler, B. L. 1999. How bacteria talk to each other: regulation of gene expression by quorum sensing. Curr. Opin. Microbiol. 2:582-587

Bergé, M. 2002. Compétence pour la transformation génétique chex la bactérie à Gram positif *Streptococcus pneumoniae*: étude du gène tardif *dprA*. Ph.D. thesis. Université Paul Sabatier de Toulouse (Sciences). Toulouse, France

Bergé, M., M. Moscoso, et al. 2002. Uptake of transforming DNA in Gram-positive bacteria: a view from *Streptococcus pneumoniae*. Mol. Microbiol. 45:411-21

Björklund, A. K., S. Light, et al. 2008. Quantitative assessment of the structural bias in protein-protein interaction assays. Proteomics 22:4657-67

Bradford, M. M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72:248-54

Cato, A. Jr., and W. R. Guild. 1968. Transformation and DNA size. I. Activity of fragments of defined size and a fit to a random double cross-over model. J. Mol. Biol. 37:157-78

Chan, P. F., K. M. O'Dwyer, et al. 2003. Characterization of a novel fucose-regulated promoter (P*fcsK*) suitable for gene essentiality and antibacterial mode-of-action studies in *Streptococcus pneumoniae*. J. Bacteriol. 185:2051-58

Chandler, M. S. 1992. The gene encoding cAMP receptor protein is required for competence development in *Haemophilus influenzae* Rd. Proc. Natl. Acad. Sci. USA 89: 1626-30

Charpentier, X., E. Kay, et al. 2011. Antibiotics and UV radiation induce competence for natural transformation in *Legionella pneumophila*. J. Bacteriol. 193:1114-21

Chen, I., and D. Dubnau. 2004. DNA uptake during bacterial transformation. Nat. Rev. Microbiol. 2: 241-9

Chen, I., R. Provvedi, et al. 2006. A macromolecular complex formed by a pilin-like protein in competent *Bacillus subtilis*. J. Biol. Chem. 281: 21720-7

Chen, J. D., and D. A. Morrison. 1987. Modulation of competence for genetic transformation in *Streptococcus pneumoniae*. J. Gen. Microbiol. 133:1959-67

Claverys, J. P., and L. S. Havarstein. 2002. Extracellular-peptide control of competence for genetic transformation in *Streptococcus pneumoniae.* Front. Biosci. 7:1798-814

Claverys, J. P., B. Martin, et al. 2007. Competence-induced fratricide in streptococci. Mol. Microbiol. 64: 1423-33

Claverys, J. P., M. Prudhomme, et al. 2006. Induction of competence regulons as a general response to stress in Gram-positive bacteria. Annu. Rev. Microbiol. 60:451–75

Claverys, J. P., M. Prudhomme, et al. 2000. Adaptation to the environment: *Streptococcus pneumoniae*, a paradigm for recombination-mediated genetic plasticity? Mol. Microbiol. 35:251-9

Dagkessamanskaia, A., M. Moscoso, et al. 2004. Interconnection of competence, stress and CiaR regulons in *Streptococcus pneumoniae*: competence triggers stationary phase autolysis of *ciaR* mutant cells. Mol. Microbiol. 51:1071-86

Dawid S., A. M. Roche, and J. N. Weiser. 2007. The blp bacteriocins of *Streptococcus pneumoniae* mediate intraspecies competition both in vitro and in vivo. Infect. Immun. 75:443-51

Desai, B. V. 2005. The induced protein CoiA promotes recombination during genetic transformation in *Streptococcus pneumonia*. PhD Thesis. University of Illinois at Chicago, Chicago, IL

Dillard, J. P. and H. S. Seifert. 2001. A variable genetic island specific for Neisseria gonorrhoeae is involved in providing DNA for natural transformation and is found more often in disseminated infection isolates. Mol. Microbiol. 41: 263-77

Dodgson, Charles Lutwidge. 1866. Alice's adventures in Wonderland. Macmillan & co., London

Dowson, C. G. 2004. What Is a Pneumococcus? p. 1-14. *In* E. I. Tuomanen, T. J. Mitchell, D. A. Morrison, and B. G. Spratt (ed.), The Pneumococcus. <u>ASM Press</u>, Washington, D.C

Dressler D. and H. Potter. 1982. Molecular mechanisms in genetic recombination. <u>Ann. Rev. Biochem.</u> 51:727-61

Duggan, S. T. 2010. Pneumococcal polysaccharide conjugate vaccine (13-valent, adsorbed) [prevenar 13®]. <u>Drugs</u> 70:1973-86

D'Souza, C., M. M. Nakano, et al. 1994. Identification of *comS*, a gene of the *srfA* operon that regulates the establishment of genetic competence in *Bacillus subtilis*. <u>Proc. Natl. Acad. Sci. USA</u> 91: 9397-401

Elkins, C, C. E. Thomas, et al. 1991. Species-specific uptake of DNA by gonococci is mediated by a 10-base-pair sequence. <u>J. Bacteriol.</u> 173:3911-3

European Antimicrobial Resistance Surveillance System (EARSS). 2008. *In. p48-55.* EARSS Annual Report in 2008.

Finkel, S. E., and R. Kolter. 2001. DNA as a nutrient: novel role for bacterial competence gene homologs. <u>J. Bacteriol.</u> 183:6288-93

Finn, E. A., M. Wolfgang, et al., 2002. Competence for natural transformation in *Neisseria gonorrhoeae*: components of DNA binding and uptake linked to type IV pilus expression. <u>Mol. Microbiol.</u> 46:749-60

Fontaine, L., D. Dandoy, et al. 2010. Development of a Versatile Procedure Based on Natural Transformation for Marker-Free Targeted Genetic Modification in *Streptococcus thermophilus*. <u>Appl. Environ. Microbiol.</u> 76:7870-7

Fox, M. S., and R. D. Hotchkiss. 1957. Initiation of bacterial transformation. <u>Nature</u> 179:1322-5

Ghosh, K., and G. D. Van Duyne. 2002. Cre-*loxP* biochemistry. <u>Methods</u> 28:374-83

Gietz, R. D., and R. A. Woods. 2002. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. <u>Methods Enzymol.</u> 350: 87-96

Gustafsson, E., P. Nilsson, et al. 2004. Characterizing the dynamics of the quorum-sensing system in *Staphylococcus aureus*. <u>J. Mol. Microbiol. Biotechnol.</u> 8:232-42

Hahn, J., and D. Dubnau. 1991. Growth stage signal transduction and the requirements for *srfA* induction in development of competence. <u>J. Bacteriol.</u> 173:7275-82

Hakenbeck, R. 1999. β-lactam resistant *Streptococcus pneumoniae*: epidemicology and evolutionary mechanism. <u>Chemotherapy</u> 45:83-94

Hakenbeck, R. 2000. Transformation in *Streptococcus pneumoniae*: mosaic genes and the regulation of competence. Res. Microbiol. 151:453-56

Hamoen, L. W, A. F. Van Werkhoven, et al. 1998. The competence transcription factor of *Bacillus subtilis* recognizes short A/T-rich **S**equences arranged in a unique, flexible pattern along the DNA helix. Genes Dev. 12:1539-50

Hamoen, L. W., W. K. Smits, et al. 2002. Improving the predictive value of the competence transcription factor (ComK) binding site in *Bacillus subtilis* using a genomic approach. Nucleic Acids Res. 30: 5517-28

Hamoen, L. W., A. F. Van Werkhoven, et al. 2000. The pleiotropic response regulator DegU functions as a priming protein in competence development in *Bacillus subtilis*. Proc. Natl. Acad. Sci. USA 97: 9246-51

Hamoen, L. W., G. Venema, et al. 2003. Controlling competence in *Bacillus subtilis*: shared use of regulators. Microbiology 149:9-17

Hare, P. D., and N. H. Chua. 2002. Excision of selectable marker genes from transgenic plants. Nat. Biotechnol. 20:575-80

Havarstein, L. S., G. Coomaraswamy, and D. A. Morrison. 1995. An unmodified heptadecapeptide pheromone induces competence for genetic transformation in *Streptococcus pneumonia*. Proc.Natl. Acad. Sci. USA 92:11140-4

Havarstein, L. S., P. Gaustad, et al. 1996. Identification of the streptococcal competence-pheromone receptor. Mol. Microbiol. 21:863-9

Håvarstein, L. S. and D. A. Morrison. 1999. Quorum-sensing and peptide pheromones in streptococcal competence for genetic transformation, p. 9-26. *In* G. M. Dunny and S. C. Winans (ed.), Cell-cell signalling in bacteria. ASM Press, Washington, D.C.

Herriott, R. M., E. M. Meyer, and M. Vogt. 1970. Defined nongrowth media for stage IIdevelopment of competence in *Haemophilus influenzae*. J. Bacteriol. 101: 517-24

Hirst, R. A., A. Kadioglu, et al. 2004. The role of pneumolysin in pneumococcal pneumonia and meningititis. Clin. Exp. Immunol. 138:195-201

Hoa, T. T., P. Tortosa, et al. 2002. Rok (YkuW) regulates genetic competence in *Bacillus subtilis* by directly repressing comK. Mol. Microbiol. 43:15-26

Hoskins, J., W. E. Alborn, Jr., et al. 2001. Genome of the bacterium *Streptococcus pneumoniae* strain R6. J. Bacteriol. 183: 5709-17

Hotchkiss, R. D. 1954. Cyclical behavior in pneumococcal growth and transformability occasioned by environmental changes. Proc. Natl. Acad. Sci. USA 40:49-55

Howden, R. 1979. A rapid bile solubility test for pneumococci. J. Clin. Pathol. 32: 1293-94

Iannelli, F., and G. Pozzi. 2004. Method for introducing specific and unmarked mutations into the chromosome of *Streptococcus pneumoniae*. Mol. Biotechnol. 26:81-6

James, P., J. Halladay, E. A. Craig.1996. Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast. Genetics 144:1425-36

Johnsborg, O.,V. Eldholm, and L. S. Håvarstein. 2007. Natural genetic transformation: prevalence, mechanisms and function. Res. Microbiol. 158:767-78

Kamiya, T.,T. Ojima, et al. 2010. Quantitative Y2H screening: cloning and signal peptide engineering of a fungal secretory LacA gene and its application to yeast two-hybrid system as a quantitative reporter. J. Biotechnol. 146:151-9

Kloosterman, T. G., J. J. Bijlsma, et al. 2006. To have neighbour's fare: extending the molecular toolbox for *Streptococcus pneumoniae*. Microbiology 152:351-59

Kreth, J., D. C. Hung, et al. 2007. The response regulator ComE in *Streptococcus mutans* functions both as a transcription activator of mutacin production and repressor of CSP biosynthesis. Microbiology 153:1799-807

Kreth, J., J. Merritt, et al. 2005. Competition and coexistence between *Streptococcus mutans* and *Streptococcus sanguinis* in the dental biofilm. J. Bacteriol. 187:7193-203

Kunst, F., M. Debarbouille, et al. 1988. Deduced polypeptides encoded by the *Bacillus subtilis sacU* locus share homology with two-component sensor-regulator systems. J. Bacteriol. 170:5093-101

Lacks, S. A. 2004. Transformation. p. 89-119. *In* E. I. Tuomanen, T. J. Mitchell, D. A. Morrison, and B. G. Spratt (ed.), The Pneumococcus. ASM Press, Washington, D.C

Lacks, S. A., and M. Neuberger. 1975. Membrane location of a deoxyribonuclease implicated inthe genetic transformation of *Diplococcus pneumoniae*. J. Bacteriol. 124:1321-9

Lambert, J. M., R. S. Bongers, and M. Kleerebezem. 2007. Cre-lox-based system for multiple gene deletions and selectable-marker removal in *Lactobacillus plantarum*. Appl. Environ. Microbiol. 73:1126-35

Lau, P. C., C. K. Sung, et al. 2002. PCR ligation mutagenesis in transformable streptococci: application and efficiency. J. Microbiol. Methods 49:193-205

Lazazzera, B. A., I. G. Kurtser, et al. 1999. An autoregulatory circuit affecting peptide signaling in *Bacillus subtilis*. J. Bacteriol. 181:5193-200

Lee, M. S. and D. A. Morrison. 1999. Identification of a new regulator in *Streptococcus pneumoniae* linking quorum sensing to competence for genetic transformation. J. Bacteriol. 181: 5004-16

Leibig, M., B. Krismer, et al. 2008. Marker removal in staphylococci via Cre recombinase and different *lox* sites. Appl. Environ. Microbiol. 74:1316-23

Lemme A., L. Gröbe, et al. 2011. Subpopulation-specific transcriptome analysis of competence-stimulating-peptide-induced *Streptococcus mutans*. J. Bacteriol. 193:1863-77

Lennette E. H., Balows A, et al. 1985. Manual of Clinical Microbiology. 4th ed. Washington, DC: American Society for Microbiology

Lorenz, M. G., and W. Wackernagel. 1994. Bacterial gene transfer by natural genetic transformation in the environment. Microbiol. Rev. 58:563-602

Luo, P. 2003. Genetic Transformation in *Streptococcus pneumoniae*: Regulation by ComX an Alternative Sigma Factor. PhD Thesis. University of Illinois at Chicago, Chicago, IL

Luo, P., H. Li, et al. 2003. ComX is a unique link between multiple quorum sensing outputs and competence in *Streptococcus pneumoniae*. Mol. Microbiol. 50: 623-33

Luo, P., H. Li, et al. 2004. Identification of ComW as a new component in the regulation of genetic transformation in *Streptococcus pneumoniae*. Mol. Microbiol. 54: 172-83

Luo, P. and D. A. Morrison. 2003. Transient association of an alternative sigma factor, ComX, with RNA polymerase during the period of competence for genetic transformation in *Streptococcus pneumoniae*. J. Bacteriol. 185: 349-58

Macfadyen, L. P. 2000. Regulation of competence development in *Haemophilus influenzae*: proposed competence regulatory elements are CRP-Binding sites. J. Theor. Biol. 207: 349-59

Marmur, J. 1961. A procedure for the isolation of DNA from microorganisms. J. Mol. Biol. 3: 11

Martin, B., M. Prudhomme, et al. 2000. Cross-regulation of competence pheromone production and export in the early control of transformation in *Streptococcus pneumoniae*. Mol. Microbiol. 38:867-78

Martin, B., Y. Quentin, et al. 2006. Independent evolution of competence regulatory cascades in streptococci? Trends Microbiol. 14: 339-45

Martinez, B, S. Skovbjerg, et al. 2009. *Streptococcus pneumoniae* autolysis prevents phagocytosis and production of phagocyte-activating cytokines. Infect. Immun. 77:3826-27

Mashburn-Warren, L., D. A. Morrison, and M. J. Federle. 2010. A novel double-tryptophan peptide pheromone controls competence in *Streptococcus* spp. via an Rgg regulator. Mol. Microbiol. 78:589-606

McIver, K. S. 2009. Stand-alone response regulators controlling global virulence networks in *Streptococcus pyogenes*. Contrib. Microbiol. 16:103-19

Mercenier, A., and B. M. Chassy. 1988. Strategies for the development of bacterial transformation systems. Biochimie 70:503-17

Méjean, V., and J. P. Claverys. 1988. Polarity of DNA entry in transformation of *Streptococcus pneumoniae*. Mol. Gen. Genet. 213:444-48

Mirouze, N., B. Martin, et al. 2007. Identification of a late gene product required to shut off expression of early com genes in *Streptococcus pneumoniae*, abstr. F-08, p.90. Abstr. 8th European Meeting on the Molecular Biology of the Pneumococcus.

Mooney, R. A., S. A. Darst, et al. 2005. Sigma and RNA polymerase: an on-again, off-again relationship? Mol. Cell 20:335-45

Morona, J. K., D. C. Miller, et al. 2004. The effect that mutations in the conserved capsular polysaccharide biosynthesis gene *cspA*, *cpsB* and *cpsD* have on virulence of *Streptococcus pneumoniae*. J. Infect. Dis. 189:1905-13

Morrison, D. A., and W. R. Guild. 1973. Breakage prior to entry of donor DNA in Pneumococcus transformation. Biochim. Biophys. Acta 299:545-56

Morrison, D. A., S. A. Lacks, et al. 1983. Isolation and characterization of three new classes of transformation-deficient mutants of *Streptococcus pneumoniae* that are defective in DNA transport and genetic recombination. J. Bacteriol. 156:281-90

Morrison, D. A., I. Mortier-Barrière, et al. 2007. Identification of the major protein component of the pneumococcal eclipse complex. J. Bacteriol. 189: 6497-500

Mortier-Barrière, I., M. Velten, et al., 2007. A key presynaptic role in transformation for a widespread bacterial protein: DprA conveys incoming ssDNA to RecA. Cell 130:824-36

Murphy, T. F., L. O. Bakaletz, and P. R. Smeesters. 2009. Microbial interactions in the respiratory tract. Pediatr. Infect. Dis. J. 28: 121-126

Ng, W. L., and B. L. Bassler. 2009. Bacterial quorum-sensing network architectures. Annu. Rev. Genet. 43:197-222

Ng, W. L., G. T. Robertson, et al. 2003. Constitutive expression of PcsB suppresses the requirement for the essential VicR (YycF) response regulator in *Streptococcus pneumoniae* R6. Mol. Microbiol. 50:1647-63

Oggioni, M. R. and D. A. Morrison. 2008. Cooperative regulation of competence development in *Streptococcus pneumoniae*: cell-to-cell signaling via a peptide pheromone and an alternative sigma factor. *In* Chemical Communication among Bacteria (S. C. Winans and B. Bassler, eds), ASM Press, Washington DC, pp. 345-362.

Ogunniyi, A. D., K. S. LeMessuier, et al. 2007. Contributions of pneumolysin, pneumococcal suface protein A (PspA) and PspC to pathogenicity of *Streptococcus pneumoniae* D39 in a mouse model. Infect. Immun. 75:1843-51

Park, I. H., D. G. Pritchard, et al. 2007. Discovery of a new capsular serotype (6C) within serogroup 6 of *Streptococcus pneumoniae*. J. Clin. Microbiol. 45:1225-33

Persuh, M., K. Turgay, et al. 1999. The N- and C-terminal domains of MecA recognize different partners in the competence molecular switch. Mol. Microbiol. 33:886-94

Pestova, E. V., L. S. Havarstein, et al. 1996. Regulation of competence for genetic transformation in *Streptococcus pneumoniae* by an auto-induced peptide pheromone and a two-component regulatory system. Mol. Microbiol. 21: 853-62

Peterkofsky, A., and C. Gazdar. 1971. Glucose and the metabolism of adenosine 3':5'-cyclic monophosphate in *Escherichia coli.* Proc. Natl. Acad. Sci. USA 68: 2794-8

Peterson, S., R. T. Cline, et al. 2000. Gene expression analysis of the *Streptococcus pneumoniae* competence regulons by use of DNA microarrays. J. Bacteriol. 182: 6192-202

Peterson, S. N., C. K. Sung, et al. 2004. Identification of competence pheromone responsive genes in *Streptococcus pneumoniae* by use of DNA microarrays. Mol. Microbiol. 51: 1051-70

Pettigrew, M. M., J. F. Gent, et al. 2008. Microbial interactions during upper respiratory tract infections. Emerg. Infect. Dis. 14:1584-91

Piotrowski, A, P. 2010. Regulation of the activities of ComX and ComW during competence development in *Streptococcus pneumoniae*. PhD Thesis. University of Illinois at Chicago, Chicago, IL

Piotrowski A, P. Luo, and D. A. Morrison. 2009 . Competence for genetic transformation in *Streptococcus pneumoniae*: termination of activity of the alternative sigma factor ComX is independent of proteolysis of ComX and ComW. J. Bacteriol. 191:3359-66

Pomerantsev, A. P., R. Sitaraman, et al. 2006. Genome engineering in *Bacillus anthracis* using Cre recombinase. Infect. Immun. 74:682-93

Prudhomme, M, et al. 2006. Antibiotic stress induces genetic transformability in the human pathogen *Streptococcus pneumoniae*. Science 313:89-92

Puyet, A., Greenberg, B., and Lacks, S.A. 1990. Genetic and structural characterization of endA: a membrane-bound nuclease required for transformation of *Streptococcus pneumoniae*. J. Mol. Biol. 213:727-38

Redfield, R. J. 2001. Do bacteria have sex? Nat. Rev. Genet. 2:634-9

Redfield, R. J., A. D. Cameron, et al. 2005. A novel CRP-dependent regulon controls expression of competence genes in *Haemophilus influenzae*. J. Mol. Biol. 347:735-47

Robertson, G. T., W. L. Ng, et al. 2002. Global transcriptional analysis of clpP mutations of type 2 *Streptococcus pneumoniae* and their effects on physiology and virulence. J. Bacteriol. 184:3508-20

Sambrook, J. and D. W. Russel. 2001. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press

Sanders, J. W., K. Leenhouts, et al. 1998. A chloride-inducible acid resistance mechanism in *Lactococcus lactis* and its regulation. Mol. Microbiol. 27: 299-310

Sauer, B. 1987. Functional expression of the cre-lox site-specific recombination system in the yeast *Saccharomyces cerevisiae*. Mol. Cell Biol. 7:2087-96

Scott, J. A. 2007. The preventable burden of pneumococcal diseases in the developing world. Vaccine 25: 2398-405

Sisco, K. L. and H. O. Smith. 1979. Sequence-specific DNA uptake in *Haemophilus* transformation. Proc. Natl. Acad. Sci. USA 76:972-6

Solomon, J. M., B. A. Lazazzera, et al. 1996. Purification and characterization of an extracellular peptide factor that affects two different developmental pathways in *Bacillus subtilis*. Genes Dev. 10: 2014-24

Standish, A. J., U. H. Stroeher, and J. C. Paton. 2005. The two-component signal transduction system RR06/HK06 regulates expression of *cbpA* in *Streptococcus pneumonia.* Proc. Natl. Acad. Sci. USA 102:7701-6

Stephen, C. W., and B. L. Bassler. Chemical communication among bacteria, Maro R. Oggroni and D.A. Morrison(ed.), Chaptor 22: 345-62

Stewart, G. J. and C. A. Carlson. 1986. The biology of natural transformation. Annu. Rev. Microbiol. 40:211-35

St-Onge, L., P. A. Furth, and P. Gruss. 1996. Temporal control of the Cre recombinase in transgenic mice by a tetracycline responsive promoter. Nucleic Acids Res. 24:3875-77

Sulavik, M. C., G. Tardif, and D. B. Clewell. 1992. Identification of a gene, rgg, which regulates expression of glucosyltransferase and influences the Spp phenotype of *Streptococcus gordonii Challis*. J. Bacteriol. 174: 3577-86

Sung, C. K. 2004. Genetic analysis in Pneumococcus: ComW as a new regulator of the competence-specific sigma factor, ComX. PhD Thesis. University of Illinois at Chicago, Chicago, IL

Sung, C. K., H. Li, et al. 2001. An rpsL cassette, janus, for gene replacement through negative selection in *Streptococcus pneumoniae*. Appl. Environ. Microbiol. 67: 5190-6

Sung, C. K., and D. A. Morrison. 2005. Two distinct functions of ComW in stabilization and activation of the alternative sigma factor ComX in *Streptococcus pneumoniae*. J. Bacteriol. 187: 3052-61

Tomasz, A., R. and D. Hotchkiss. 1964. Regulation of the transformability of peumococcal cultures by macromolecular cell products. Proc. Natl. Acad. Sci. USA 51:480-7.

Tomasz, A. 1997. Antibiotic resistance in *Streptococcus pneumoniae*. Clin. Infect. Dis. 24:85-8

Tomb, J. F., H. Amitai, and S. Abuchakra. 1993. A competence-specific regulatory sequence of *Haemophilus influenzae* (Abstract). Molecular Genetics of Bacteria and Phages, Cold Spring Harbor Laboratory Press

Tran, T. D., H. Y. Kwon, et al. 2011. Decrease in Penicillin Susceptibility Due to Heat Shock Protein ClpL in *Streptococcus pneumonia*e. Antimicrob. Agents Chemother. 55:2714-28

Trappetti, C., L. Guadi, et al. 2011. The impact of the competence quorum sensing system on *Streptococcus pneumoniae* biofilms varies depending on the experimental model. BMC Microbiol. 11: 75

Turgay, K., J. Hahn, et al. 1998. Competence in *Bacillus subtilis* is controlled by regulated proteolysis of a transcription factor. Embo. J. 17: 6730-8

Vagner, V., J. P. Claverys, et al. 1990. Direction of DNA entry in competent cells of *Bacillus subtilis*. Mol. Microbiol. 4:1785-8

van der Ploeg, J. R. 2005. Regulation of bacteriocin production in *Streptococcus mutans* by the quorum-sensing system required for development of genetic competence. J. Bacteriol. 187:3980-9

van Sinderen, D., A. Luttinger, et al. 1995. *comK* encodes the competence transcription factor, the key regulatory protein for competence development in *Bacillus subtilis*. Mol. Microbiol. 15: 455-62

Yan, X., H. J. Yu, et al. 2008. Cre/lox System and PCR-Based Genome Engineering in *Bacillus subtilis*. Appl. Environ. Microbiol. 74: 5556-62

Ween, O., P. Gaustad, et al. 1999. Identification of DNA binding sites for ComE, a key regulator of natural competence in *Streptococcus pneumoniae*. Mol. Microbiol. 33: 817-27

White, B. 1938. The biology of pneumococcus, Second printing, 1979 edition (Cambridge, MA, Harvard University Press)

Zhang, Z., and B. Lutz. 2002. Cre recombinase-mediated inversion using *lox66* and *lox71*: method to introduce conditional point mutations into the CREB-binding protein. Nucleic Acids Res. 30:90

Zuo, J., Q. W. Niu, et al. 2001. Chemical-regulated, site-specific DNA excision in transgenic plants. Nat. Biotechnol. 19:157-61

APPENDICES

Appendix A

<u>Investigation of the interaction between ComX and ComW with surface plasmon resonance</u>

   To further verify the interaction between ComX and ComW in vitro, I sought to purify these proteins and explore their interaction via surface plasmon resonance (SPR). ComX was purified with E. coli strain BL21::pXPL01, in which a copy of comX was fused with His6-tag and induced by the promoter of lac operon. After the induction of IPTG, the ComX derivative was expressed and purified via nickel column. The final concentration is 0.2 mg/ml. However, ComW has been successfully purified alone because of its insolubility, even though several purification system had been tried. It seems to us that ComW is a pretty hydrophobic protein. It might be part of some unknown protein complex. Therefore, instead of using ComW to explore interaction with ComX, a fusion protein, MBP-ComW, was used as substitute to improve its solubility. In pMALTM protein fusion system (New England Biolabs), a copy of comW was fused to the C-terminal of MBP to make expressing plasmid pAP01, which is induced by the promoter of lac opern. The E. coli strain (BL21::pAP01) was then grown and induce with IPTG. The lysate was mounted onto the amylose resin column and purified. The final concentration of the purified MBP-ComW was 2.5 mg/ml. Meanwhile, MBP protein alone was used as the negative control. So the SPR signal between ComX and MBP-ComW was compared to the signal between ComX and MBP. In Biacore T100 instrument, 50µg of ComX was used to immobilize ComX onto the CM5 censor chip as the ligand. The analytes, MBP-ComW and MBP alone, were passed over the chip surface individually. The sensorgram was monitored and analyzed. Unfortunately, both MBP-ComW and MBP alone did not show binding affinity with ComX on the censor chip. This could be due to the facts that the interacting face of ComW was blocked by MBP from access, or the CM5 chip did not work out for ComX.

Appendix B

Investigation of post-translational modification of ComX by mass spectrum

Even though the above SPR experiment did not provide positive result, I am still confident about interaction between ComX and ComW from yeast two-hybrid assay. I went further to hypothesize that probably it is this interaction that activates ComX to function. And this activation could be any forms of post-translational modifications, such as phosphorylation, methylation and ubiquitination. To test this hypothesis, I intended to use mass spectrum to see if there is any structural difference between ComX expressed in *comW* proficient and deficient background. Two strains were constructed through a serial of transformations: CP2156 (*comX::his₆, ssbB::lacZ, ΔclpP*) and CP2157 (*comX::his₆, ssbB::lacZ, ΔclpP, ΔcomW*). ComX was purified from the two strains after CSP induction for 15 min at OD_{550}=0.1 and submitted for PTM analysis by mass spectrometry. It is expected that there might be some structure difference in the ComX protein produced in *comW* wildtype and mutant backgrounds. If a mass difference is verified, I want to go further to figure what kind of modification was made and on what residue(s).

Appendix C

S. pneumoniae strains constructed in this study
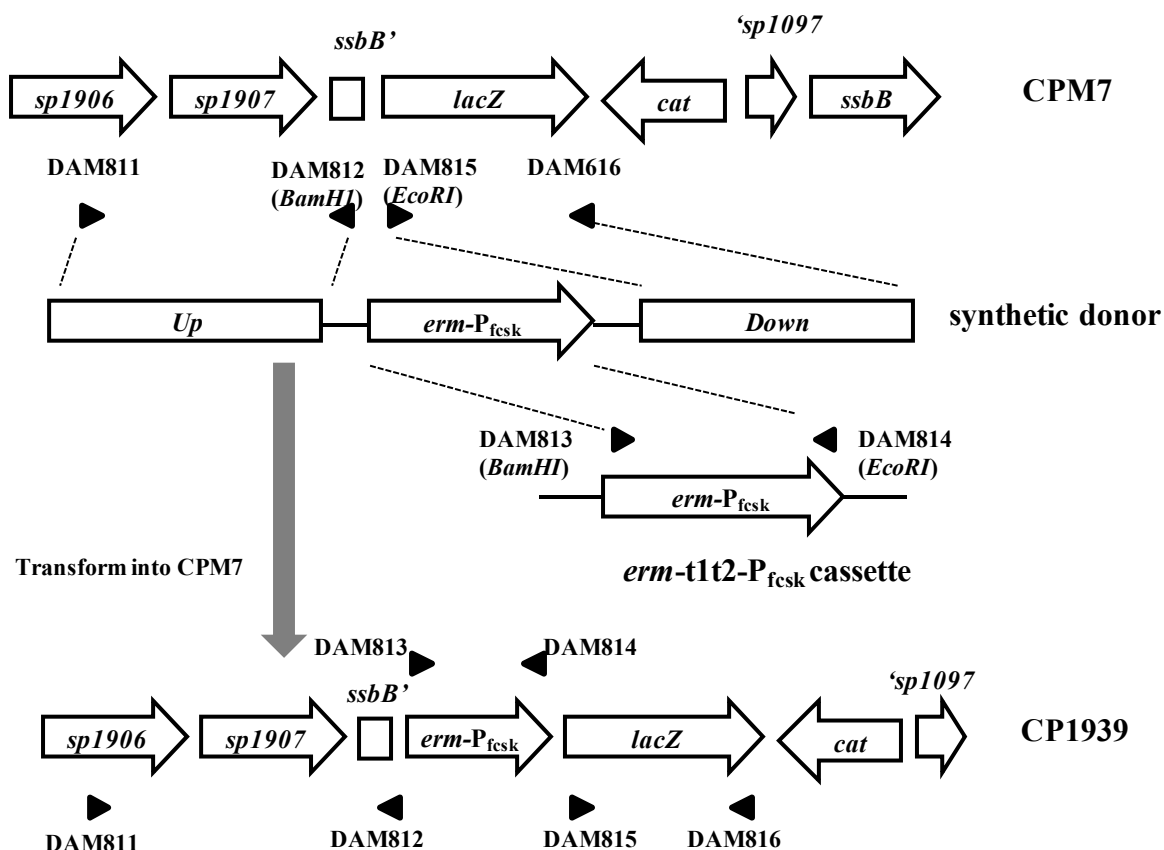
Appendix C (Continued)

Figure 24 Insertion of *erm*-t1t2-P$_{fcsk}$ into CPM7 to make CP1939. To insert *erm*-t1t2-P$_{fcsk}$ cassette in front of *lacZ* resporter, an upstream sequence, encompassing part of *ssbB*, *sp1907*, and part of *sp1906,* was amplified with primer pair DAM811/812 from genomic DNA of CPM7 and digested with BamHI endonuclease. A downstream sequence, encompassing part of *lacZ*, was amplified with primer pair DAM815/816 from genomic DNA of CPM7 and digested with EcoRI endonuclease. The *erm*-t1t2-P$_{fcsk}$ cassette was amplified with primer pair DAM813/814, provided by Gregory Roberston, and double-digested with BamHI and EcoRI. The three fragments were ligated together and transformed into CPM7 to make the strain CP1939. The construction was verified by sequencing PCR products amplified with DAM811/814 and DAM813/816.
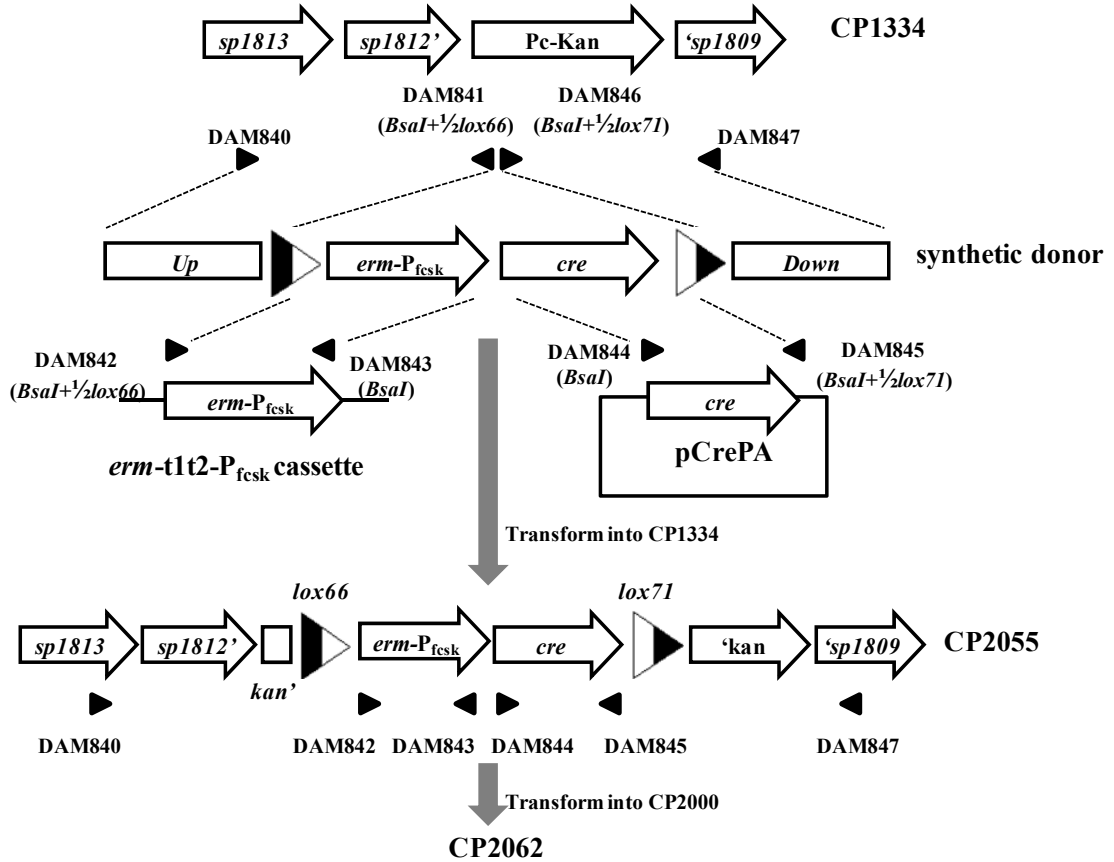
Figure 25 Construction of Cheshire cassette (CP2055 and CP2062). To create *lox66*/*erm*-P$_{fcsk}$/*cre*/*lox71* for the application of markerless deletion, four fragments were amplified. An upstream sequence, encompassing part of sp1813, sp1812, and kan, was amplified with primer pair DAM840/841 from genomic DNA of CP1334 and digested with BsaI endonuclease. An *erm*-P$_{fcsk}$ cassette was amplified with primer pair DAM842/843 from *erm*-t1t2-P$_{fcsk}$ cassette and digested with *BsaI*. A *cre* gene was amplified with primer pair DAM844/845 from plamid pCrePA and digested with *BsaI*. A downstream sequence, encompassing part of *kan*, was amplified with primer pair DAM846/847 from genomic DNA of cp1334 and digested with BsaI. A *lox66* site was spitted into two halves, one half attached to primer DAM841 and the other attached to DAM842. A *lox71* site was spitted into two halves, one half attached to primer DAM845 and the other attached to DAM846. The four fragments were ligated together and transformed into CP1334 to make the strain CP2055. The genomic DNA of CP2055 was transformed into CP2000 to make CP2062. The construction was verified by sequencing PCR products amplified with DAM840/843, DAM842/845, and DAM844/847.

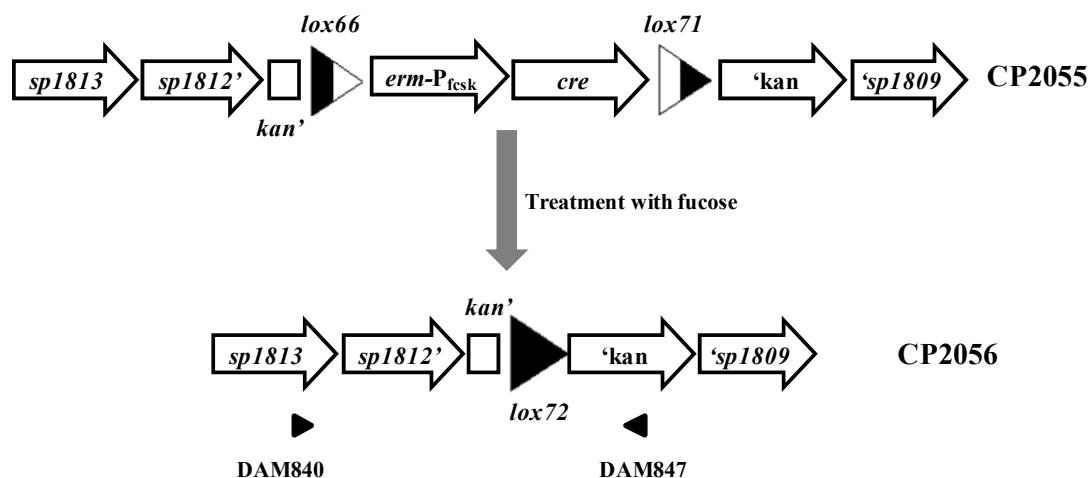**Appendix C (Continued)**

Figure 26 Construction of strain CP2056 (Pc-Kan::*lox72*). To obtain kanamycin-resistant CP2056, a culture of CP2055 was growing in complete CAT medium plus 0.1% fucose. When cell density reached about OD_{550}=0.8, the cells were plated onto compete CAT agar medium plus 200 µg/ml kanamycin, with the methods described in (**2.6.1**). The Kan^R clones were CP2056, verified by sequencing the PCR product amplified by DAM840/847.
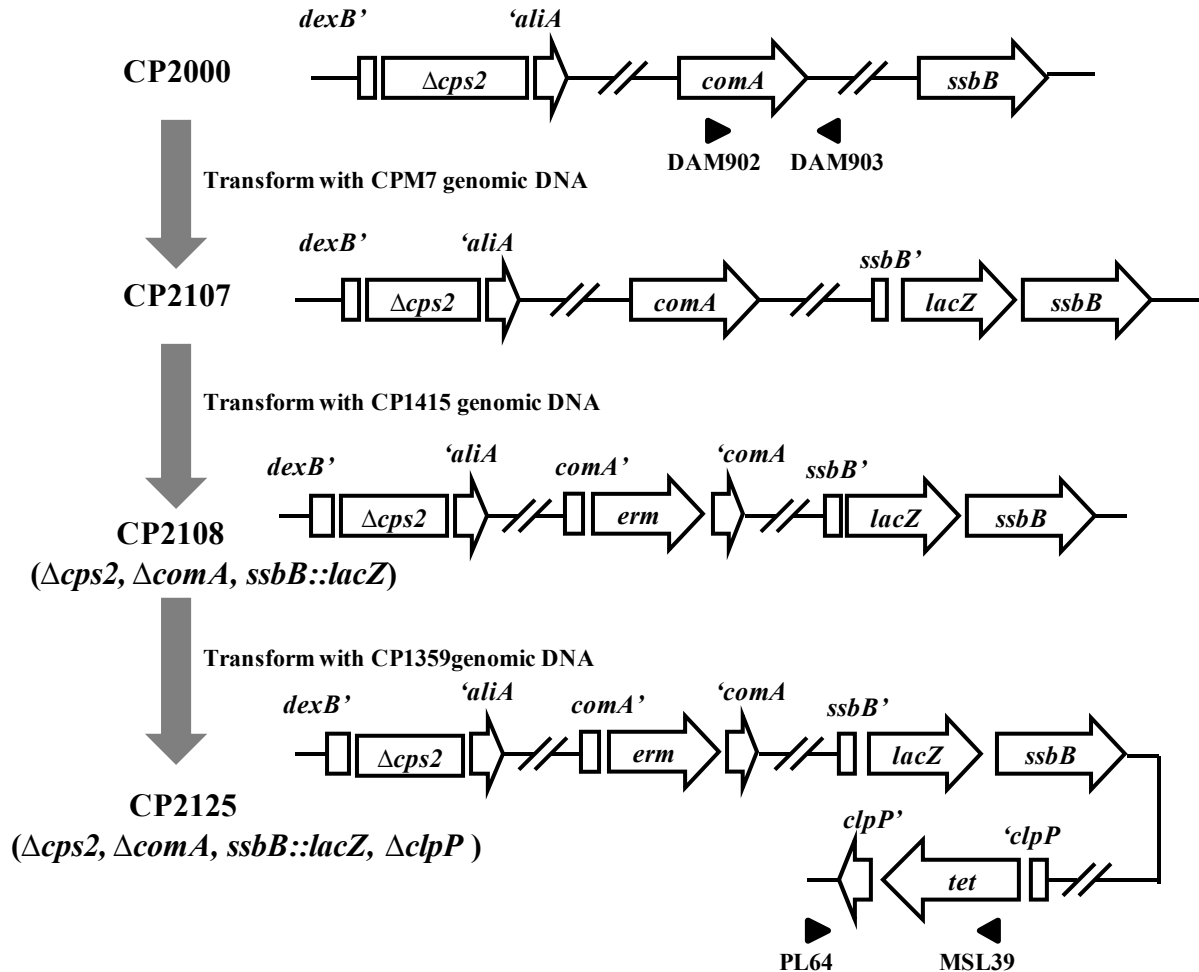
Appendix C (Continued)

Figure 27 Construction of parental strains, CP2108 and CP2125. To create the parental strains for making late gene mutants, CP2000 was first transformed with CPM7 (*ssbB*::*lacZ*) genomic DNA to put a lacZ reporter in *ssbB* operon. The intermediating strain, CP2107, was subsequently transformed with CP1415 (Δ*comA*) genomic DNA to incorporate *comA* deletion. Eventually, the parental strain CP2108 was obtained. Another parental strain, CP2125, was obtained by transforming CP2108 with the genomic DNA of CP1359 (Δ*clpP*). The two parental strains were verified by checking the loss of PCR product amplified by DAM902/903, by checking the gain of LacZ activity, and by checking the gain of PCR product amplified PL64/MSL39.
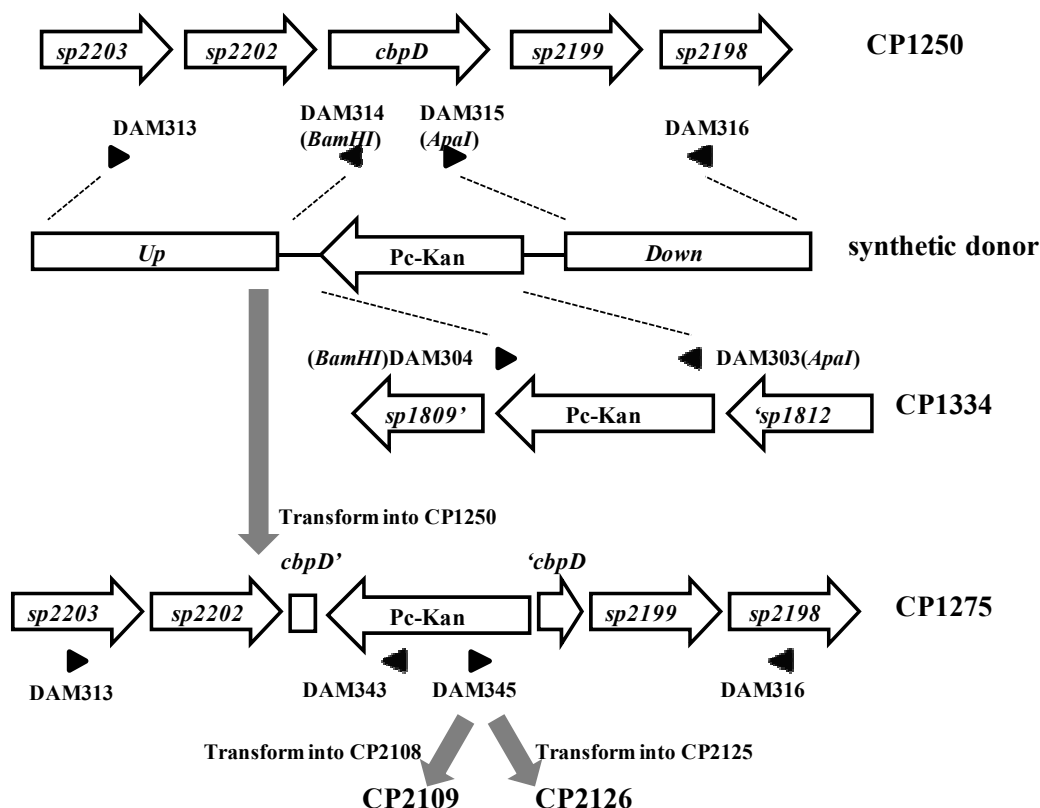
Appendix C (Continued)

Figure 28 Derivation of strains CP2109 and CP2126 (Δ*cbpD*::Pc-Kan). To create Δ*cbpD* strains, an upstream sequence, encompassing part of *sp2203*, *sp2202*, and part of *cbpD,* was amplified with primer pair DAM313/314 from genomic DNA of CP1250 and digested with BamHI endonuclease. A downstream sequence, encompassing part of *cbpD*, *sp2199*, and part of *sp2198*, was amplified with primer pair DAM315/316 from genomic DNA of CP1250 and digested with ApaI endonuclease. A kanamycin-resistant marker, Pc-Kan, was amplified with primer pair DAM303/304 from genomic DNA of CP1334 and double-digested with BamHI and ApaI. The three fragments were ligated together and transformed into CP1250 to make the strain CP1275 (Sung, C. K., 2004). Genomic DNA of CP1275 was used as donor to transform strains CP2108 and CP2125 to make CP2109 and CP2126, respectively. The deletion was verified by checking the gain of PCR products amplified with DAM313/343 and DAM345/316.
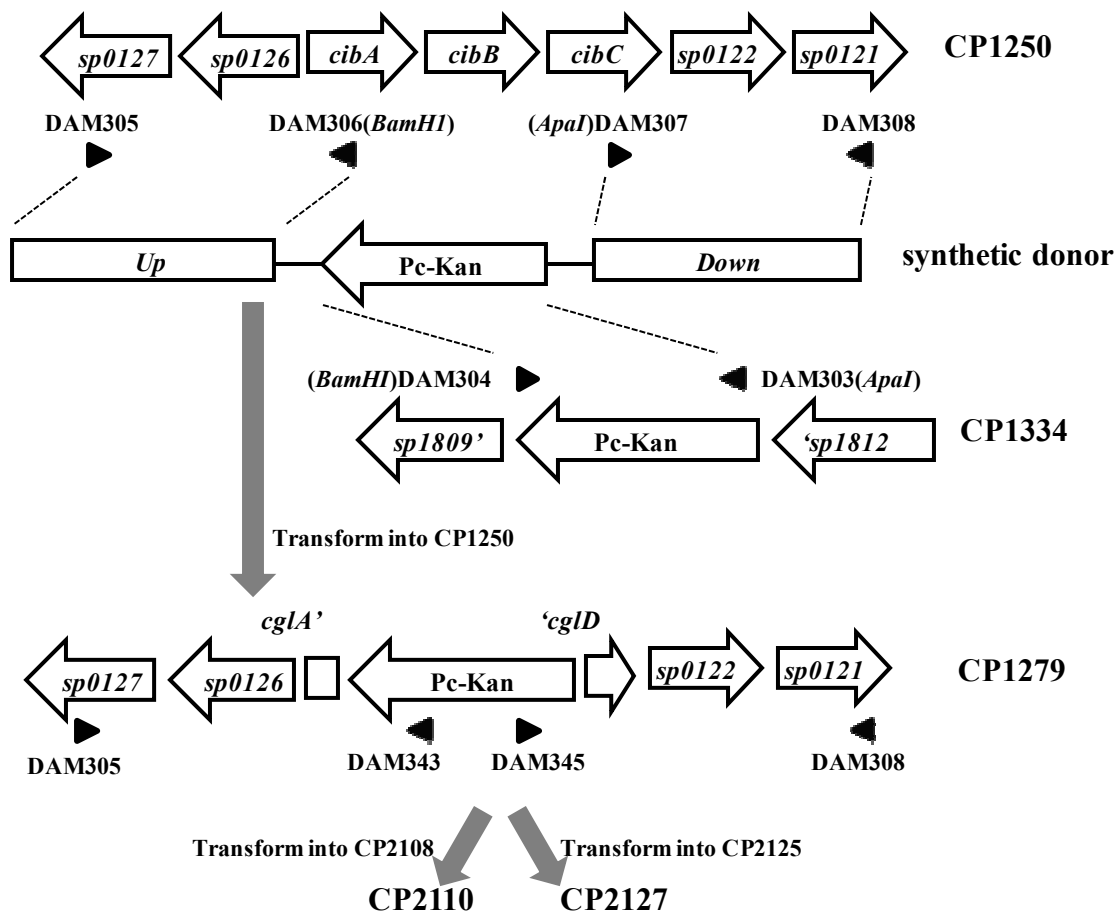
Appendix C (Continued)

Figure 29 Derivation of strains CP2110 and CP2127 (Δ*cibABC*::Pc-Kan). To create Δ*cibABC* strains, an upstream sequence, encompassing part of *sp0127*, *sp0126*, and part of *cibA,* was amplified with primer pair DAM305/306 from genomic DNA of CP1250 and digested with BamHI endonuclease. A downstream sequence, encompassing part of *cibC*, *sp0122*, and part of *sp0121*, was amplified with primer pair DAM307/308 from genomic DNA of CP1250 and digested with ApaI endonuclease. A kanamycin-resistant marker, Pc-Kan, was amplified with primer pair DAM303/304 from genomic DNA of CP1334 and double-digested with BamHI and ApaI. The three fragments were ligated together and transformed into CP1250 to make the strain CP1279 (Sung, C. K., 2004). Genomic DNA of CP1279 was used as donor to transform strains CP2108 and CP2125 to make CP2110 and CP2127, respectively. The deletion was verified by checking the gain of PCR products amplified with DAM305/343 and DAM308/345.
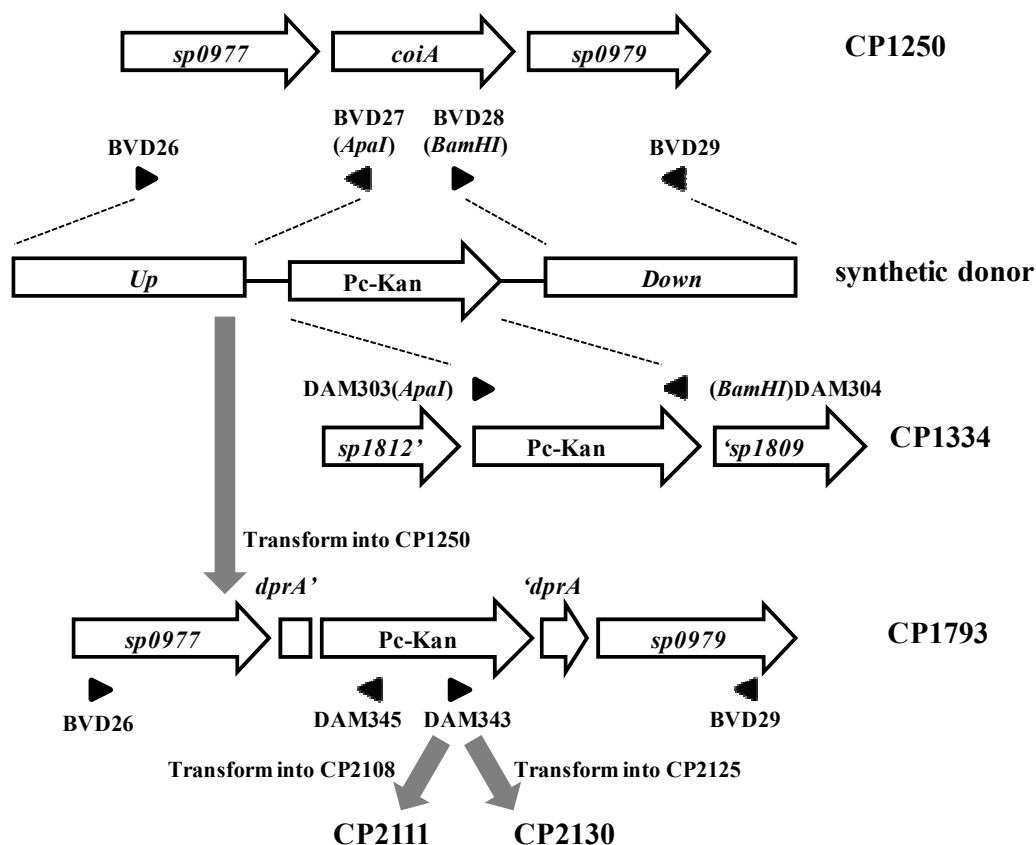
Appendix C (Continued)

Figure 30 Derivation of strains CP2111 and CP2130 (Δ*coiA*::Pc-Kan). To create Δ*coiA* strains, an upstream sequence, encompassing part of *sp0977* and part of *dprA*, was amplified with primer pair BVD26/27 from genomic DNA of CP1250 and digested with ApaI endonuclease. A downstream sequence, encompassing part of *dprA* and part of *sp0979*, was amplified with primer pair BCD28/29 from genomic DNA of CP1250 and digested with BamHI endonuclease. A kanamycin-resistant marker, Pc-Kan, was amplified with primer pair DAM303/304 from genomic DNA of CP1334 and double-digested with BamHI and ApaI. The three fragments were ligated together and transformed into CP1250 to make the strain CP1793 (Desai, B. V., 2005). Genomic DNA of CP1793 was used as donor to transform strains CP2108 and CP2125 to make CP2111 and CP2130, respectively. The deletion was verified by checking the gain of PCR products amplified with BVD26/DAM345 and DAM343/BVD29.
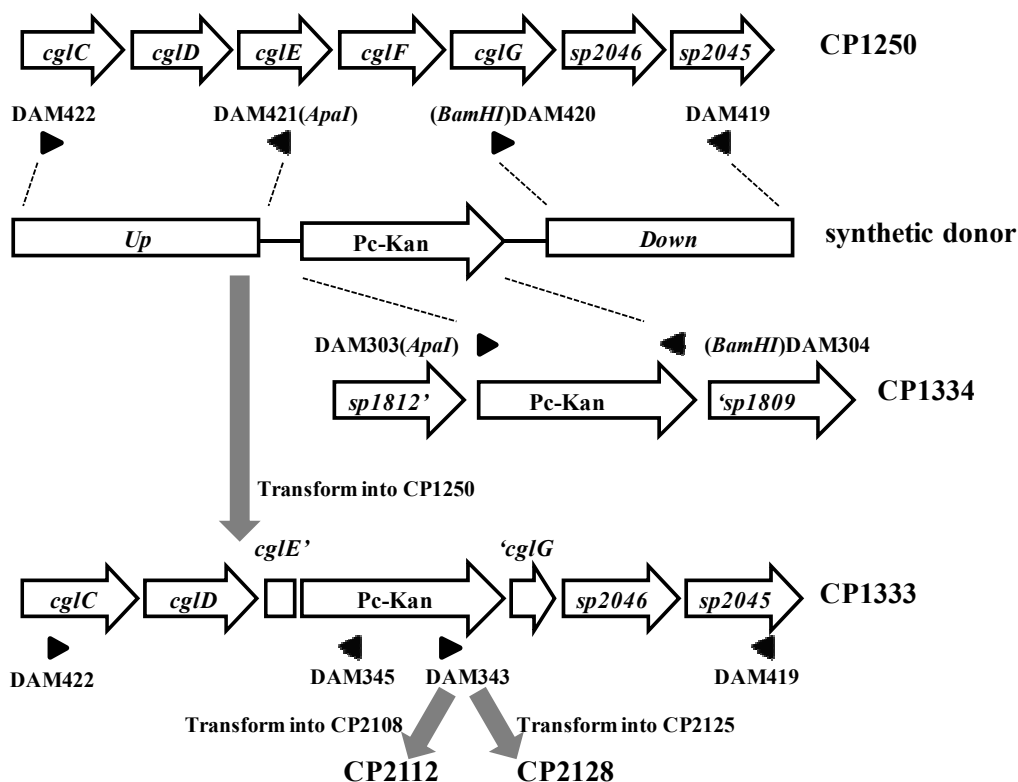
Appendix C (Continued)

Figure 31 Derivation of strains CP2112 and CP2128 (ΔcglEFG::Pc-Kan). To create ΔcglEFG strains, an upstream sequence, encompassing part of *clgC*, *cglD*, and part of *cglE,* was amplified with primer pair DAM422/421 from genomic DNA of CP1250 and digested with ApaI endonuclease. A downstream sequence, encompassing part of *cglG, sp2046*, and part of *sp2045*, was amplified with primer pair DAM419/420 from genomic DNA of CP1250 and digested with BamHI endonuclease. A kanamycin-resistant marker, Pc-Kan, was amplified with primer pair DAM303/304 from genomic DNA of CP1334 and double-digested with BamHI and ApaI. The three fragments were ligated together and transformed into CP1250 to make the strain CP1333 (Sung, C. K., 2004). Genomic DNA of CP1333 was used as donor to transform strains CP2108 and CP2125 to make CP2112 and CP2128, respectively. The deletion was verified by checking the gain of PCR products amplified with DA422/345 and DAM343/419, and by a transformation deficiency.
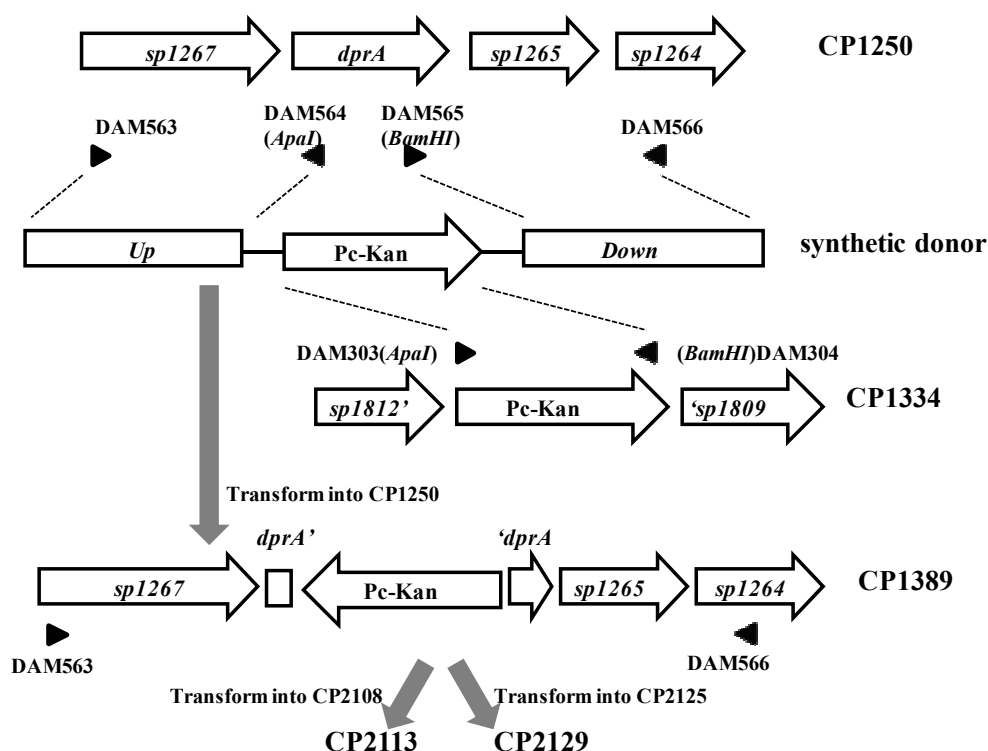
Figure 32 Derivation of strains CP2113 and CP2129 (Δ*dprA*::Pc-Kan). To create Δ*dprA* strains, an upstream sequence, encompassing part of *sp1267* and part of *dprA,* was amplified with primer pair DAM563/564 from genomic DNA of CP1250 and digested with ApaI endonuclease. A downstream sequence, encompassing part of *dprA, sp1265*, and part of *sp1264*, was amplified with primer pair DAM565/566 from genomic DNA of CP1250 and digested with BamHI endonuclease. A kanamycin-resistant marker, Pc-Kan, was amplified with primer pair DAM303/304 from genomic DNA of CP1334 and double-digested with BamHI and ApaI. The three fragments were ligated together and transformed into CP1250 to make the strain CP1389 (Sung, C. K., 2004). Genomic DNA of CP1389 was used as donor DNA to transform strains CP2108 and CP2125 to make CP2113 and CP2129, respectively. The deletion was verified by checking the size change of PCR products amplified with DAM563/566, and a transformation deficiency.
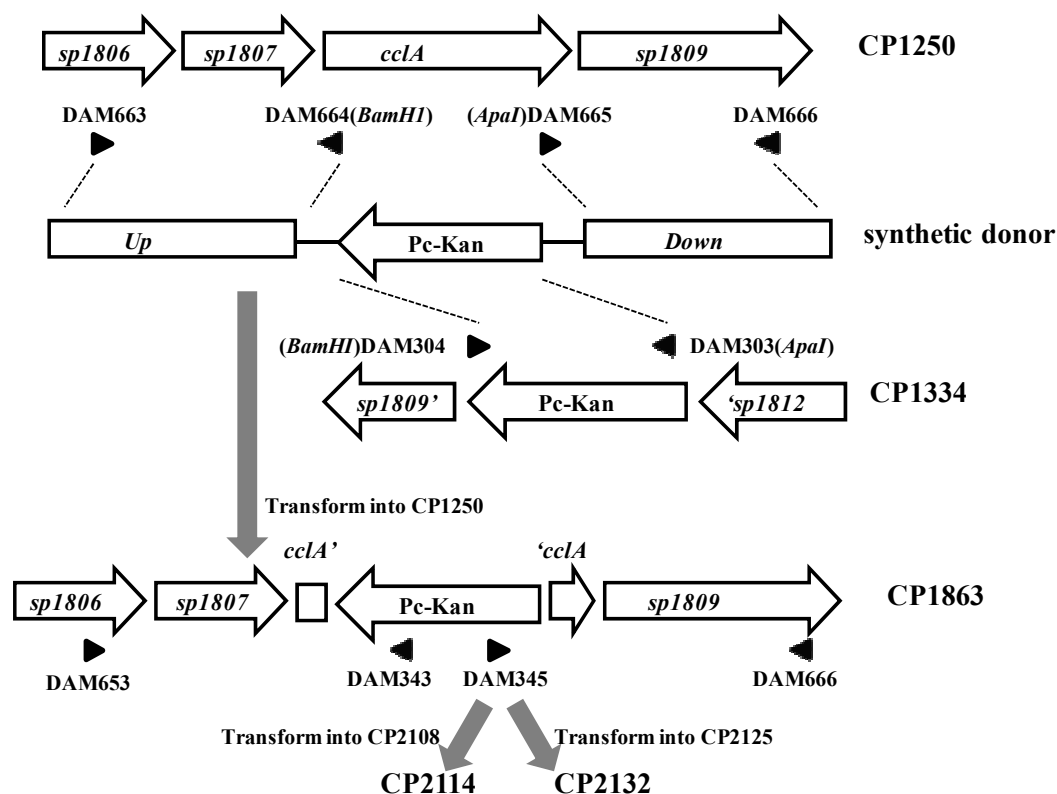
Appendix C (Continued)

Figure 33 Derivation of strains CP2114 and CP2132 (ΔcelAB ::Pc-Kan). To create ΔcclA strains, an upstream sequence, encompassing part of *sp1806*, *sp1807*, and part of *cclA,* was amplified with primer pair DAM663/664 from genomic DNA of CP1250 and digested with BamHI endonuclease. A downstream sequence, encompassing part of *cclA*, and part of *sp1809*, was amplified with primer pair DAM661/662 from genomic DNA of CP1250 and digested with ApaI endonuclease. A kanamycin-resistant marker, Pc-Kan, was amplified with primer pair DAM303/304 from genomic DNA of CP1334 and double-digested with BamHI and ApaI. The three fragments were ligated together and transformed into CP1250 to make the strain CP1863. Genomic DNA of CP1863 was then used as donor to transform strains CP2108 and CP2125 to make CP2114 and CP2132, respectively. The deletion was verified by checking the gain of PCR products amplified with DAM663/343 and DAM345/666, and by a transformation deficiency.
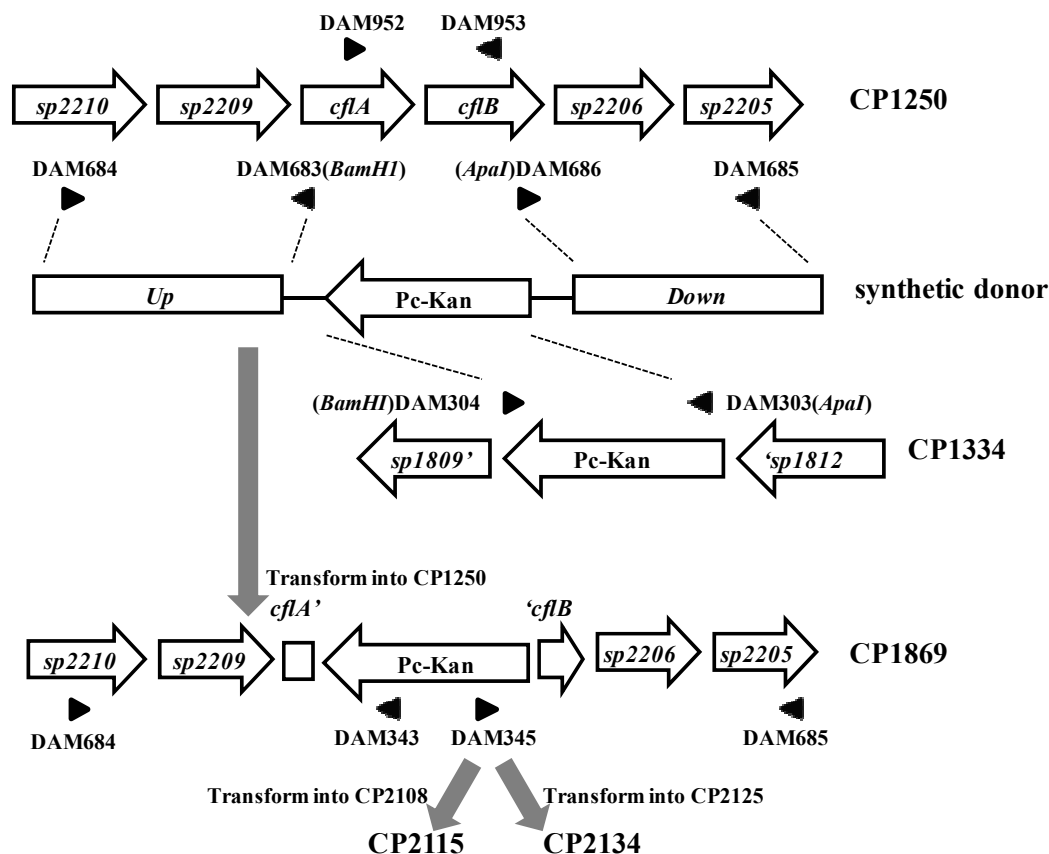
Appendix C (Continued)

Figure 34 Construction of CP2115 and CP2134 (Δ*cflAB*::Pc-Kan). To create Δ*cflAB* strains, an upstream sequence, encompassing part of *sp2210*, *sp2209*, and part of *cflA,* was amplified with primer pair DAM684/683 from genomic DNA of CP1250 and digested with BamHI endonuclease. A downstream sequence, encompassing part of *cflB*, *sp2206*, and part of *sp2207*, was amplified with primer pair DAM685/686 from genomic DNA of CP1250 and digested with ApaI endonuclease. A kanamycin-resistant marker, Pc-Kan, was amplified with primer pair DAM303/304 from genomic DNA of CP1334 and double-digested with BamHI and ApaI. The three fragments were ligated together and transformed into CP1250 to make the strain CP1869. Genomic DNA of CP1869 was then used as donor to transform strains CP2108 and CP2125 to make CP2115 and CP2134, respectively. The deletion was verified by checking the gain of PCR products amplified with DA684/343 and DAM685/345, by checking the loss of PCR product amplified by DAM952/953, and by a transformation deficiency.

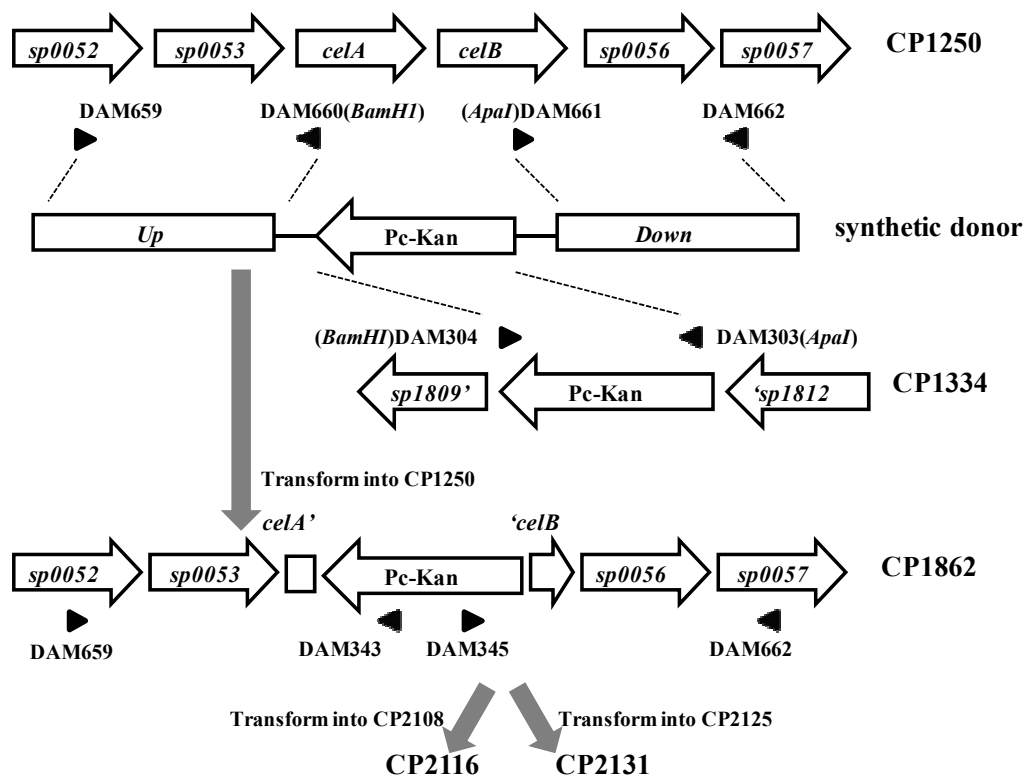**Figure 35 Construction of strains CP2116 and CP2131 (Δ*celAB*::Pc-Kan)**. To create Δ*celAB* strains, an upstream sequence, encompassing part of *sp0052*, *sp0053*, and part of *celA,* was amplified with primer pair DAM659/660 from genomic DNA of CP1250 and digested with BamHI endonuclease. A downstream sequence, encompassing part of *celB, sp0056*, and part of *sp0057*, was amplified with primer pair DAM661/662 from genomic DNA of CP1250 and digested with ApaI endonuclease. A kanamycin-resistant marker, Pc-Kan, was amplified with primer pair DAM303/304 from genomic DNA of CP1334 and double-digested with BamHI and ApaI. The three fragments were ligated together and transformed into CP1250 to make the strain CP1862. Genomic DNA of CP1862 was then used as donor DNA to transform strains CP2108 and CP2125 to make CP2116 and CP2131, respectively. The deletion was verified by checking the gain of PCR products amplified with DAM659/343 and DAM345/662, and by a transformation deficiency.
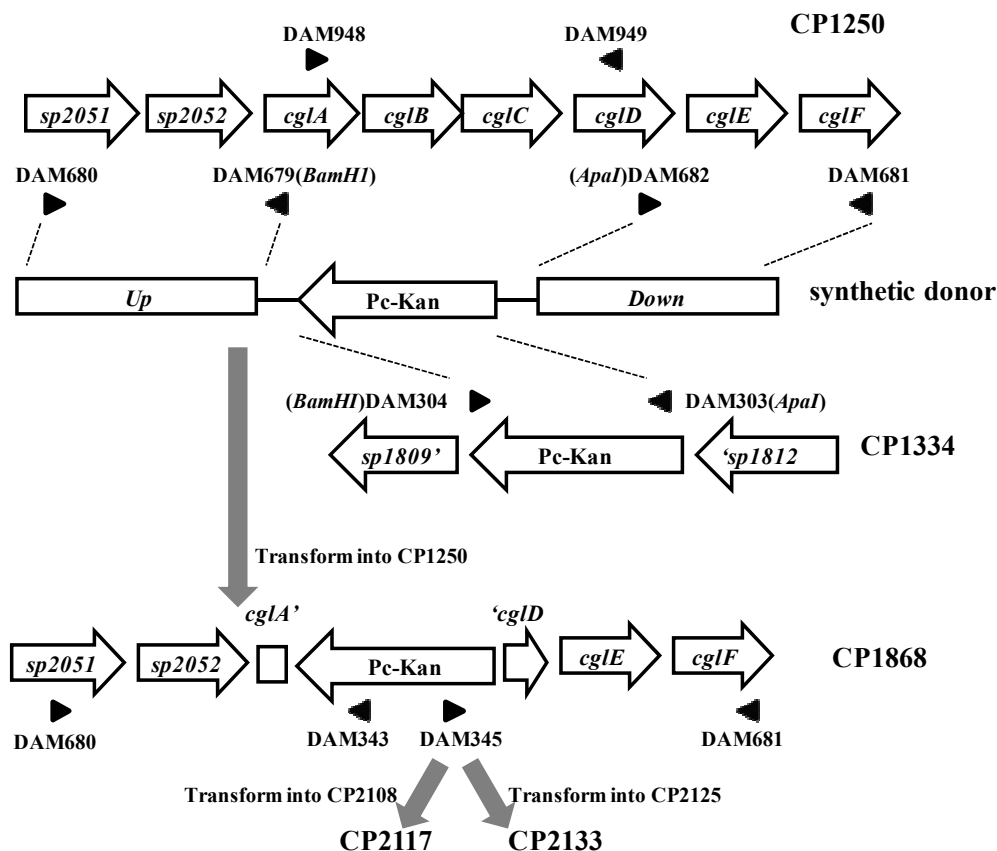
Appendix C (Continued)

Figure 36 Construction of strains CP2117 and CP2133 (Δ*cglABCD*::Pc-Kan). To create Δ*cglABCD* strains, an upstream sequence, encompassing part of *sp2051*, *sp2052*, and part of *cglA,* was amplified with primer pair DAM680/679 from genomic DNA of CP1250 and digested with BamHI endonuclease. A downstream sequence, encompassing part of *cglD*, *cglE*, and part of *cglF*, was amplified with primer pair DAM682/681 from genomic DNA of CP1250 and digested with ApaI endonuclease. A kanamycin-resistant marker, Pc-Kan, was amplified with primer pair DAM303/304 from genomic DNA of CP1334 and double-digested with BamHI and ApaI. The three fragments were ligated together and transformed into CP1250 to make the strain CP1868. Genomic DNA of CP1868 was then used as donor to transform strains CP2108 and CP2125 to make CP2117 and CP2133, respectively. The deletion was verified by checking the gain of PCR products amplified with DA680/343 and DAM681/345, by checking the loss of PCR product amplified by DAM948/949, and by a transformation deficiency.
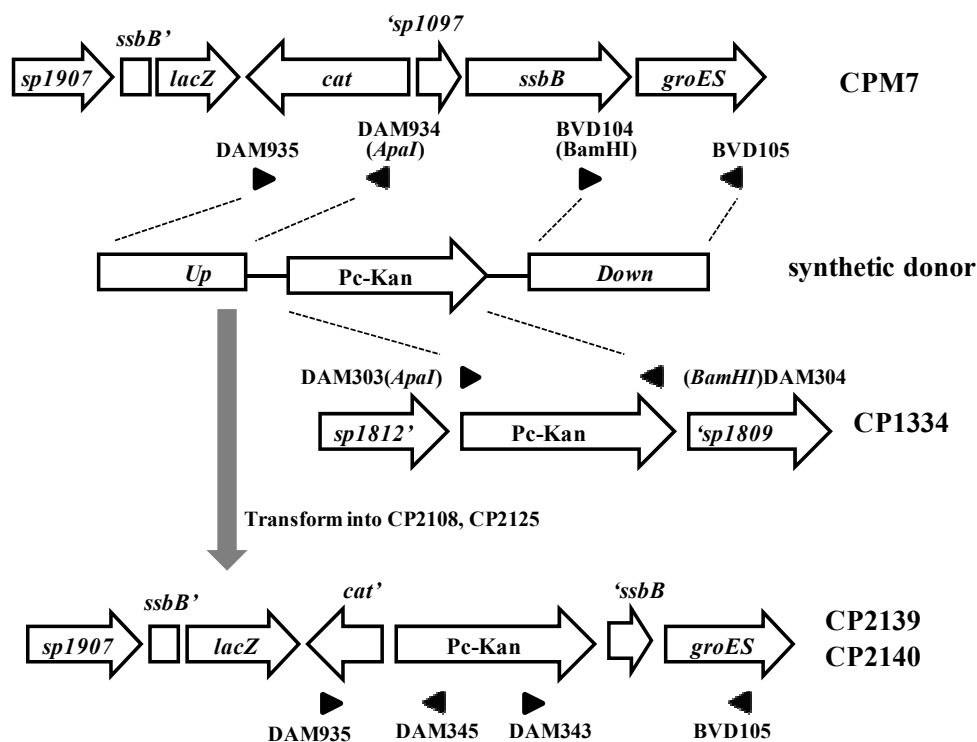
Appendix C (Continued)

Figure 37 Construction of strains CP2139 and CP2140 (Δ*ssbB*::Pc-Kan). To create Δ*ssbB* strains, an upstream sequence, encompassing part of *cat*, was amplified with primer pair DAM935/934 from genomic DNA of CPM7 and digested with ApaI endonuclease. A downstream sequence, encompassing part of *ssbB* and part of *groES*, was amplified with primer pair BVD104/105 from genomic DNA of CPM7 and digested with BamHI endonuclease. A kanamycin-resistant marker, Pc-Kan, was amplified with primer pair DAM303/304 from genomic DNA of CP1334 and double-digested with BamHI and ApaI. The three fragments were ligated together and transformed into CP2108 and CP2125 to make the strains CP2139 and CP2140, respectively. The deletion was verified by checking the gain of PCR products amplified with DA935/345 and DAM343/BVD105, and by a transformation deficiency.
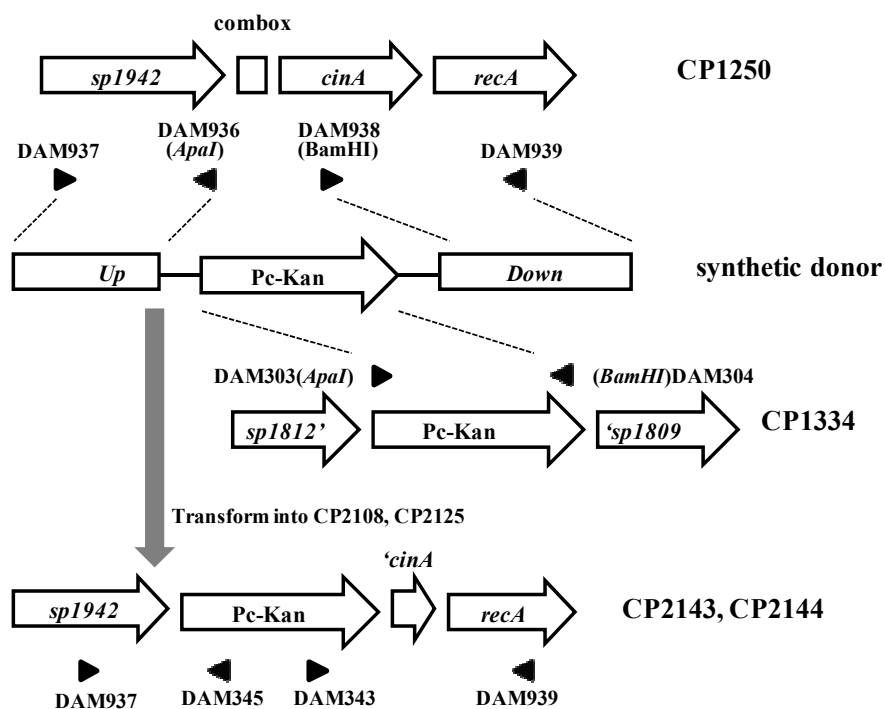
Appendix C (Continued)

Figure 38 Construction of strains CP2143 and CP2144 (ΔPcom-*cinA*::Pc-Kan). To create ΔPcom-*cinA* strains, an upstream sequence, encompassing part of *sp1942*, was amplified with primer pair DAM936/937 from genomic DNA of cp1250 and digested with ApaI endonuclease. A downstream sequence, encompassing part of *cinA* and part of *recA*, was amplified with primer pair DAM938/939 from genomic DNA of CP1250 and digested with BamHI endonuclease. A kanamycin-resistant marker, Pc-Kan, was amplified with primer pair DAM303/304 from genomic DNA of CP1334 and double-digested with BamHI and ApaI. The three fragments were ligated together and transformed into CP2108 and CP2125 to make the strains CP2143 and CP2144, respectively. The deletion was verified by checking the gain of PCR products amplified with DAM937/345 and DAM343/DAM939.
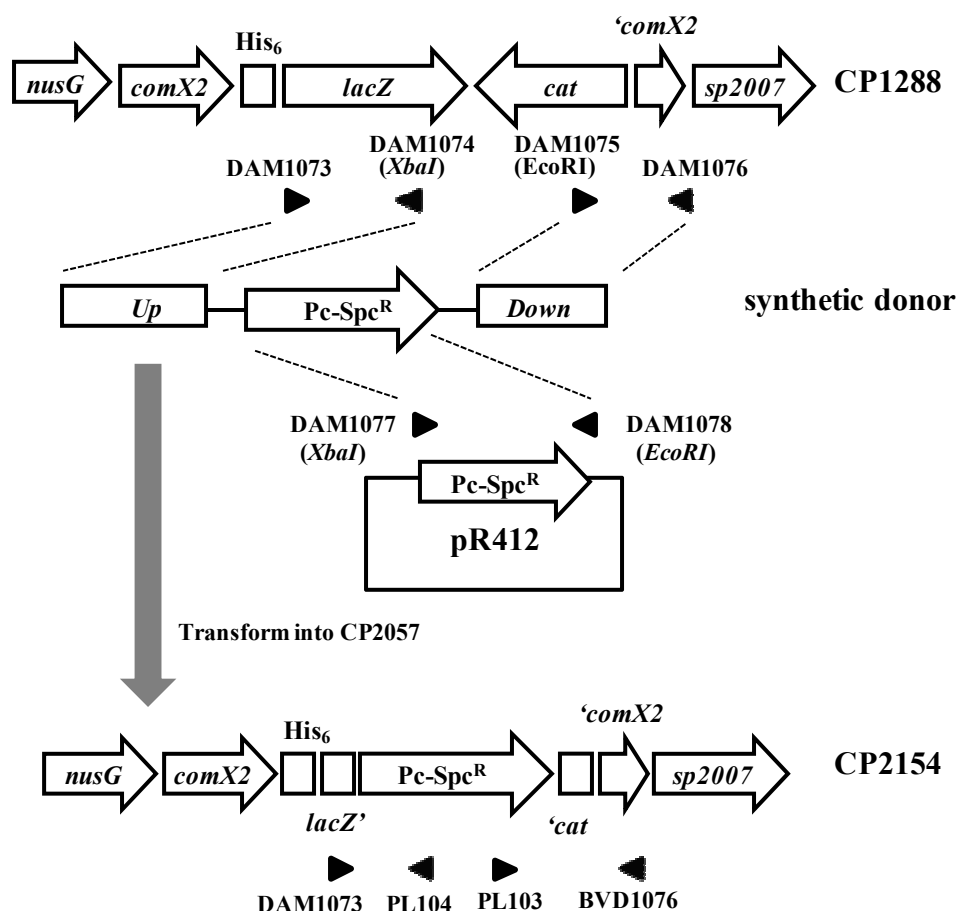
Figure 39 Construction of strain CP2154 (ΔpEVP3::Pc-Kan). To disrupt pEVP3 cassette in strain CP2057, an upstream sequence, encompassing part of *lacZ*, was amplified with primer pair DAM1073/1074 from genomic DNA of CP1288 and digested with XbaI endonuclease. A downstream sequence, encompassing part of *cat* and part of *comX2*, was amplified with primer pair DAM1075/1076 from genomic DNA of CP1288 and digested with EcoRI endonuclease. A spectinomycin-resistant marker, Pc-Spc, was amplified with primer pair DAM1077/1078 from plasmid DNA of pR412 and double-digested with XbaI and EcoRI. The three fragments were ligated together and transformed into CP2057 to make the strains CP2154. The deletion was verified by checking the gain of PCR products amplified with DAM1073/PL104 and DAM1076/PL103, and by checking the loss of LacZ activity.

Appendix D

Construction of plasmids used in this study.

Appendix D (Continued)

Figure 40 Map of the *E. coli* and yeast shuttle plasmid pACT2 used for constructions in yeast two-hybrid assay. The plasmid carries the DNA activation domain of Gal4 transcription factor, whose expression is driven by ADH1 promoter and stopped by ADH1 terminator. The *E. coli* and yeast replication origins make the plasmid viable in both species. An ampicillin-resistant marker is used for the selection in *E. coli*, while a leucine marker is for the selection in yeast. The target protein is usually fused to the C-terminus of the activation domain of Gal4, by inserting from the multiple cloning sites (*NcoI, SfiI, XmaI, SmaI, BamHI, EcoRI, SacI, EcoICRI, BanII, XhoI,* and *BglII*). In my design, the *BamHI* and *XhoI* sites were used.

Appendix D (Continued)
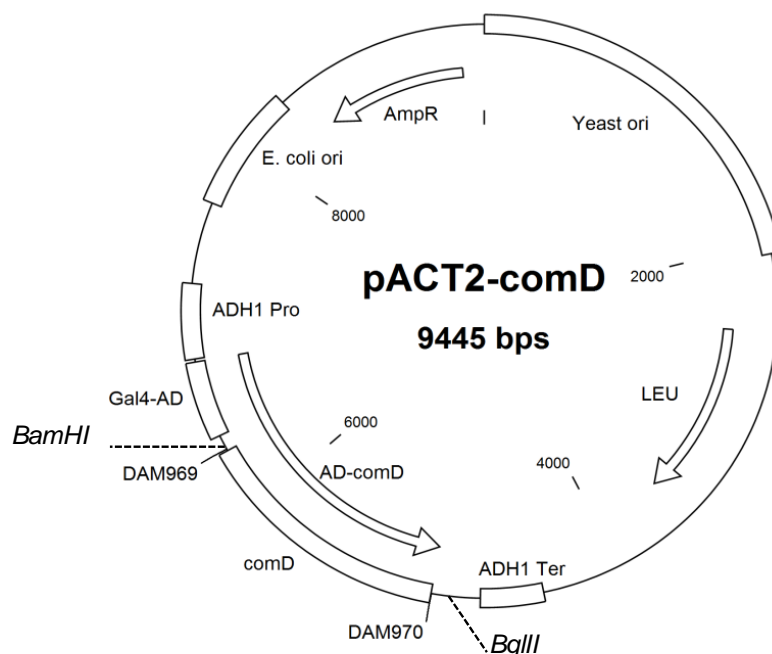
Linker region between Gal4-AD and *comD*:

```
6301  cccaaatcca aataaatcca ttactctttg gatccccggg gcctccatgg ccatatgacc
      gggtttaggt ttatttaggt aatgagaaac ctaggggccc cggaggtacc ggtatactgg
      <........comD........<<
        g  f  g  f  l  d  m
      <...........................AD-comD...........................<
        g  f  g  f  l  d  m  v  r  q  i  g  p  a  e  m  a  m  h  g

6361  acccaagcta gcgtaatctg gaacatcgta tgggtaagcc atacagatct cttttttttgg
      tgggttcgat cgcattagac cttgtagcat acccattcgg tatgtctaga gaaaaaaacc
                                                        <<.Gal4-AD..<
                                                          e  k  k  p
      <...........................AD-comD...........................<
        g  l  s  a  y  d  p  v  d  y  p  y  a  m  c  i  e  k  k  p
```

Figure 41 Construction of chimeric plasmid pACT2-*comD*. To insert *comD* into plasmid pACT2, a copy of the *comD* was amplified with primer pair DAM969/970 and double-digested with *BamHI* and *XhoI* endonucleases. Meanwhile, circular pACT2 plasmid DNA was also double-digested with *BamHI* and *XhoI* into linear vector. The insert and vector were then ligated together to make the chimeric plasmid, pATC2-*comD*. Translation of the linker region showed that *comD* was inserted in the right ORF. The presence of *comD* in the chimeric plasmid was verified by enzyme digestion (BamHI and BglII), as shown in Fig. 15. The insert was cut off by BamHI, indicating that the ligation was accurate.

Appendix D (Continued)
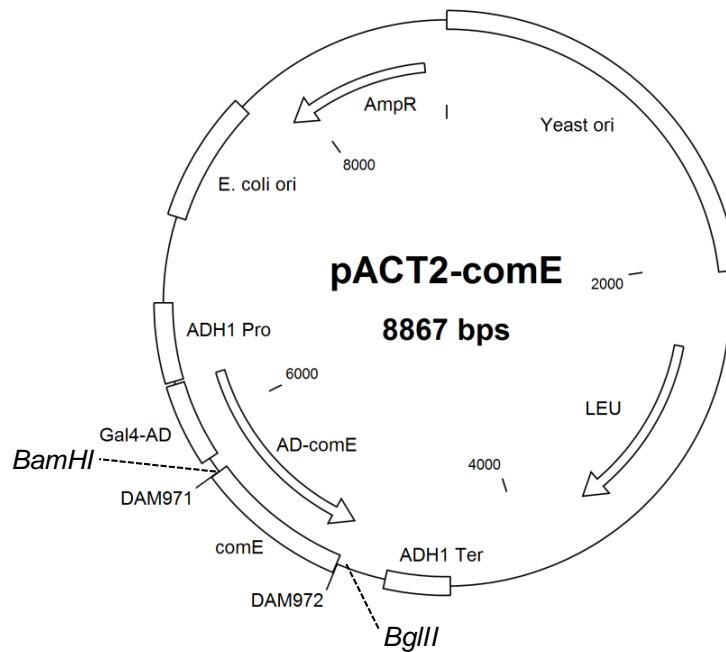
Linker region between Gal4-AD and *comE*:

```
5741   aactttcatt cggatccccg gggcctccat ggccatatga ccacccaagc tagcgtaatc tggaacatcg
       ttgaaagtaa gcctaggggc cccggaggta ccggtatact ggtgggttcg atcgcattag accttgtagc
       <.comE.<<
         v  k  m
       <..................................AD-comE.................................<
         v  k  m  r  i  g   p  a  e  m   a  m  h   g  g  l   s  a  y  d   p  v  d
5811   tatgggtaag ccatacagat ctcttttttt gggtttggtg gggtatcttc atcatcgaat agatagttat
       atacccattc ggtatgtcta gagaaaaaaa cccaaaccac cccatagaag tagtagctta tctatcaata
                            <<.....................Gal4-AD.....................<
                              e  k  k   p  n  p   p  t  d  e   d  d  f   l  y  n
       <.............................AD-comE.................................<
         y  p  y   a  m  c  i   e  k  k   p  n  p   p  t  d  e   d  d  f   l  y  n
```

Figure 42 Construction of chimeric plasmid pACT2-*comE*. To insert *comE* into plasmid pACT2, a copy of the comE was amplified with primer pair DAM971/972 and double-digested with *BamHI* and *XhoI* endonucleases. Meanwhile, circular pACT2 plasmid DNA was also double-digested with *BamHI* and *XhoI* into linear vector. The insert and vector were then ligated together to make the chimeric plasmid, pATC2-*comE*. Translation of the linker region showed that *comE* was inserted in the right ORF. The presence of *comE* in the chimeric plasmid was verified by enzyme digestion (*BamHI* and *BglII*), as shown in Fig. 15. The insert was cut off by BamHI, indicating that the ligation was accurate.

Appendix D (Continued)
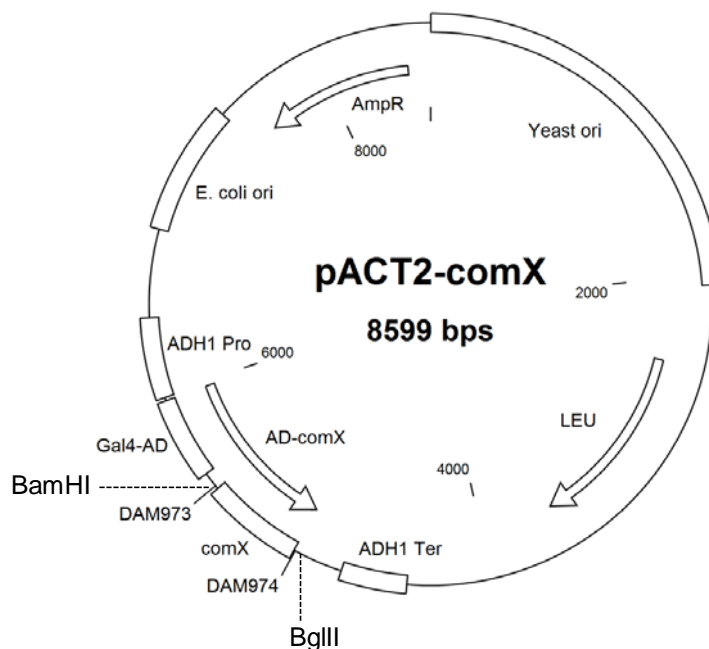

Linker region between Gal4-AD and *comX*:

```
5461   ttctttaatc ataattttcc cctggatccc cggggcctcc atggccatat gaccacccaa
       aagaaattag tattaaaagg ggacctaggg gccccggagg taccggtata ctggtgggtt
       <...comX...<<
         e  k  i    m
       <..........................AD-comX...........................<
         e  k  i    m  i  k    g  q  i  g    p  a  e    m  a  m    h  g  g  l

5521   gctagcgtaa tctggaacat cgtatgggta agccatacag atctcttttt ttgggtttgg
       cgatcgcatt agaccttgta gcatacccat tcggtatgtc tagagaaaaa aacccaaacc
                                                   <<....Gal4-AD.....<
                                                     e  k  k  p  n  p
       <..........................AD-comX...........................<
         s  a  y    d  p  v    d  y  p  y    a  m  c    i  e  k    k  p  n  p
```

Figure 43 Construction of chimeric plasmid pACT2-*comX*. To insert *comX* into plasmid pACT2, a copy of the *comX* was amplified with primer pair DAM973/974 and double-digested with *BamHI* and *XhoI* endonucleases. Meanwhile, circular pACT2 plasmid DNA was also double-digested with *BamHI* and *XhoI* into linear vector. The insert and vector were then ligated together to make the chimeric plasmid, pATC2-*comX*. Translation of the linker region showed that *comX* was inserted in the right ORF. The presence of *comX* in the chimeric plasmid was verified by enzyme digestion (*BamHI* and *BglII*), as shown in Fig. 15. The insert was cut off by BamHI, indicating that the ligation was accurate.

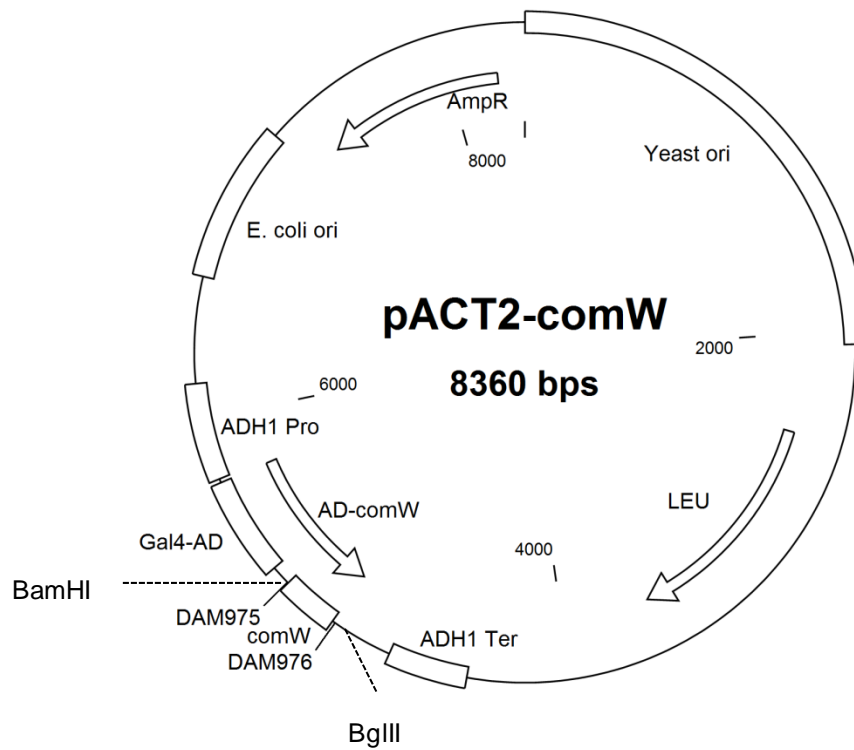**Appendix D (Continued)**

Linker region between Gal4-AD and *comW*:

```
5221   gctcataaat tttttgtaac ataaggatcc ccggggcctc catggccata tgaccaccca
       cgagtattta aaaaacattg tattcctagg ggccccggag gtaccggtat actggtgggt
       <.........comW........<<
       q  e  y  i   k  q  l   m
       <...........................AD-comW...........................<
       q  e  y  i   k  q  l   m  l  i   g  p  a  e   m  a  m   h  g  g

5281   agctagcgta atctggaaca tcgtatgggt aagccataca gatctctttt tttgggtttg
       tcgatcgcat tagaccttgt agcataccca ttcggtatgt ctagagaaaa aaacccaaac
                                                   <<...Gal4-AD.....<
                                                     e  k   k  p  n
       <...........................AD-comW...........................<
       l  s  a  y   d  p  v   d  y  p   y  a  m  c   i  e  k   k  p  n
```

Figure 44 Construction of chimeric plasmid pACT2-*comW*. To insert *comW* into plasmid pACT2, a copy of the *comW* was amplified with primer pair DAM975/976 and double-digested with *BamHI* and *XhoI* endonucleases. Meanwhile, circular pACT2 plasmid DNA was also double-digested with *BamHI* and *XhoI* into linear vector. The insert and vector were then ligated together to make the chimeric plasmid, pATC2-*comW*. Translation of the linker region showed that *comW* was inserted in the right ORF. The presence of *comW* in the chimeric plasmid was verified by enzyme digestion (*BamHI* and *BglII*), as shown in Fig. 15. The insert was cut off by BamHI,indicating that the ligation was accurate.

Appendix D (Continued)
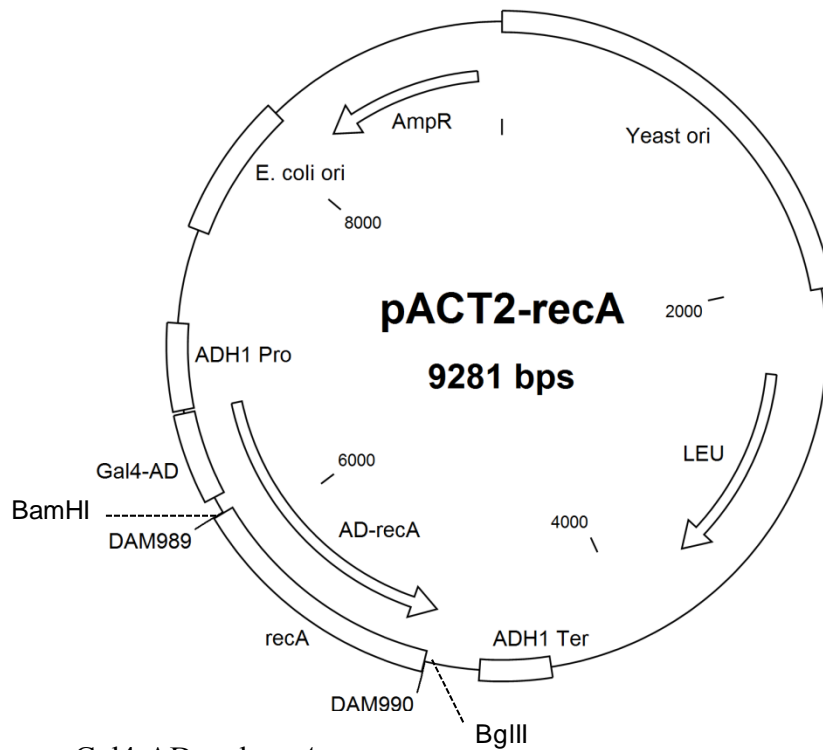
Linker region between Gal4-AD and *recA*:

```
6121  attttttga aatttcttct aatttttttg gtttttttcgc cattcggatc cccggggcct ccatggccat atgaccaccc aagctagcgt
      taaaaaaact ttaaagaaga ttaaaaaaac caaaaaagcg gtaagcctag gggccccgga ggtaccggta tactggtggg ttcgatcgca
      <....................recA....................<<
      f  k  k  s  i  e  e  l  k  k  p  k  k  a  m
      <.................................AD-recA.........................................<
      f  k  k  s  i  e  e  l  k  k  p  k  k  a  m  r  i  g  p  a  e  m  a  m  h  g  g  l  s  a

6211  aatctggaac atcgtatggg taagccatac agatctcttt ttttgggttt ggtggggtat cttcatcatc gaatagatag ttatatacat
      ttagaccttg tagcataccc attcggtatg tctagagaaa aaaacccaaa ccaccccata gaagtagtag cttatctatc aatatatgta
                                     <<........................Gal4-AD........................<
                                      e  k  p  n  p  p  t  d  e  d  d  f  l  y  n  y  v
      <..............................AD-recA...............................................<
      y  d  p  v  d  y  p  y  a  m  c  i  e  k  p  n  p  p  t  d  e  d  d  f  l  y  n  y  v
```

Figure 45 Construction of chimeric plasmid pACT2-*recA*. To insert *recA* into plasmid pACT2, a copy of the *recA* was amplified with primer pair DAM989/990 and double-digested with *BamHI* and *XhoI* endonucleases. Meanwhile, circular pACT2 plasmid DNA was also double-digested with *BamHI* and *XhoI* into linear vector. The insert and vector were then ligated together to make the chimeric plasmid, pATC2-*recA*. Translation of the linker region showed that *recA* was inserted in the right ORF. The presence of *recA* in the chimeric plasmid was verified by enzyme digestion (*BamHI* and *BglII*), as shown in Fig. 15. The insert was cut off by BamHI, indicating that the ligation was accurate.
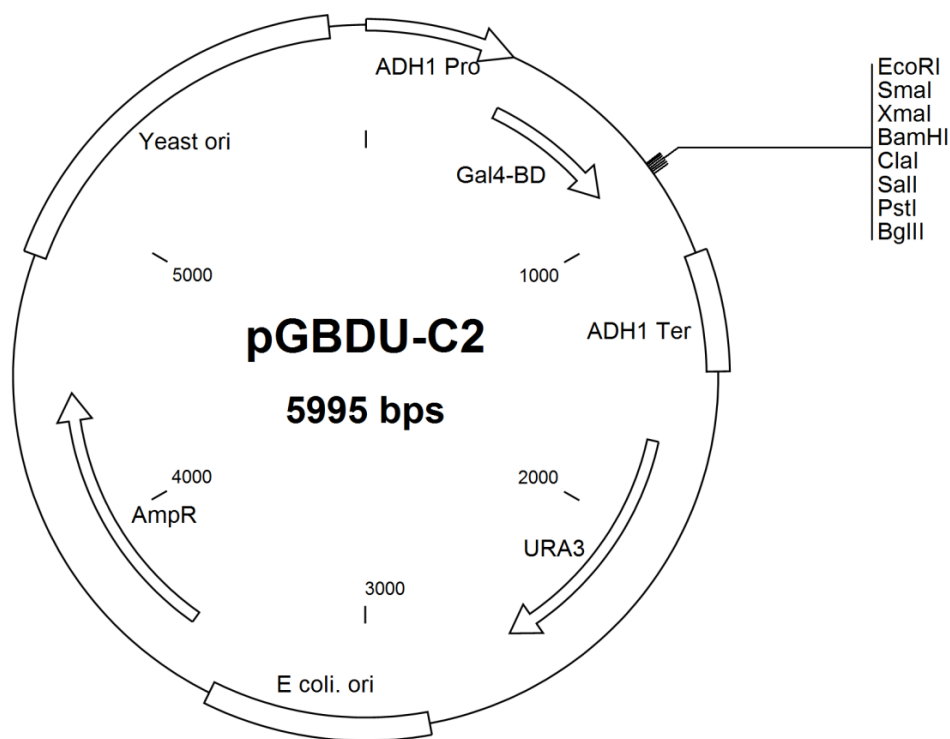
Appendix D (Continued)

Figure 46 Map of the *E. coli* and yeast shuttle plasmid pGBDUC2 used for constructions in yeast two-hybrid assay. The plasmid carries the DNA binding domain of Gal4 transcription factor, whose expression is driven by ADH1 promoter and stopped by ADH1 terminator. The *E. coli* and yeast replication origins make the plasmid viable in both species. An ampicillin-resistant marker is used for the selection in *E. coli*, while a uracil marker is for the selection in yeast. The target protein is usually fused to the C-terminus of the binding domain of Gal4, by inserting from the multiple cloning sites (EcoRI, SmaI, XmaI, BamHI, ClaI, SacI, PstI, and BglII). In my design, the BamHI and SalI sites were used.

Appendix D (Continued)

Linker region of Gal4-BD and *dprA*:

```
841   tagtaacaaa ggtcaaagac agttgactgt atcgccggaa ttcccgggga tccagttatt
      atcattgttt ccagtttctg tcaactgaca tagcggcctt aagggcccct aggtcaataa
      >..............Gal4-BD..............>>
      s   s   n   k   g   q   r   q   l   t   v   s
      >...........................BD-dprA...........................>
      s   s   n   k   g   q   r   q   l   t   v   s   p   e   f   p   g   i   q   l

901   tatgaaaatc acaaactatg aaatctataa gttaaaaaaa tcaggtttga ccaatcaaca
      atactttag tgtttgatac tttagatatt caattttttt agtccaaact ggttagttgt
      >...........................BD-dprA...........................>
      f   m   k   i   t   n   y   e   i   y   k   l   k   k   s   g   l   t   n   q
       >>.........................DrpA.............................>
          m   k   i   t   n   y   e   i   y   k   l   k   k   s   g   l   t   n   q
```

Figure 47 Construction of chimeric plasmid pGBDUC2-*dprA*. To insert *dprA* into plasmid pGBDUC2, a copy of the *dprA* was amplified with primer pair DAM967/968 and double-digested with BamHI and salI endonucleases. Meanwhile, circular pGBDUC2 plasmid DNA was also double-digested with BamHI and salI into linear vector. The insert and vector were then ligated together to make the chimeric plasmid, pGBDUC2-*dprA*. Translation of the linker region showed that *dprA* was inserted in the right ORF. The presence of *dprA* in the chimeric plasmid was verified by enzyme digestion (BamHI and BglII), as shown in Fig. 15. The insert was cut off by BamHI suggested that the ligation was accurate.
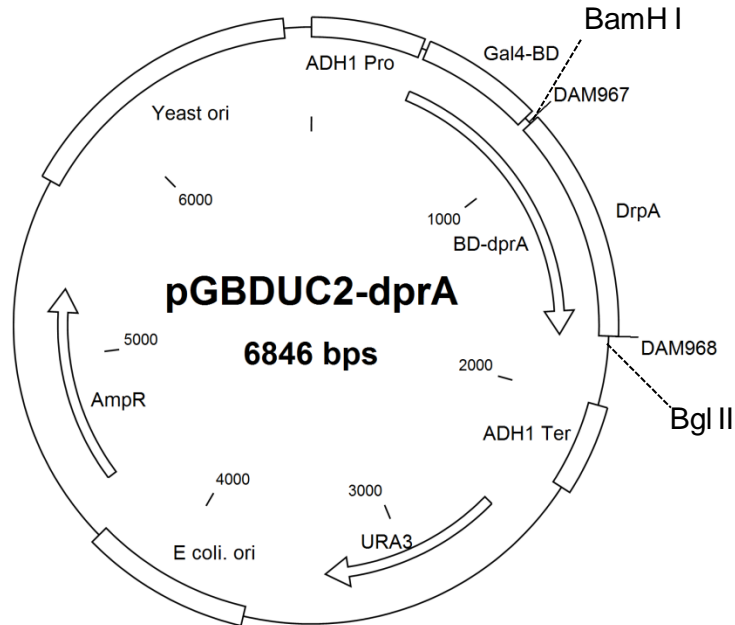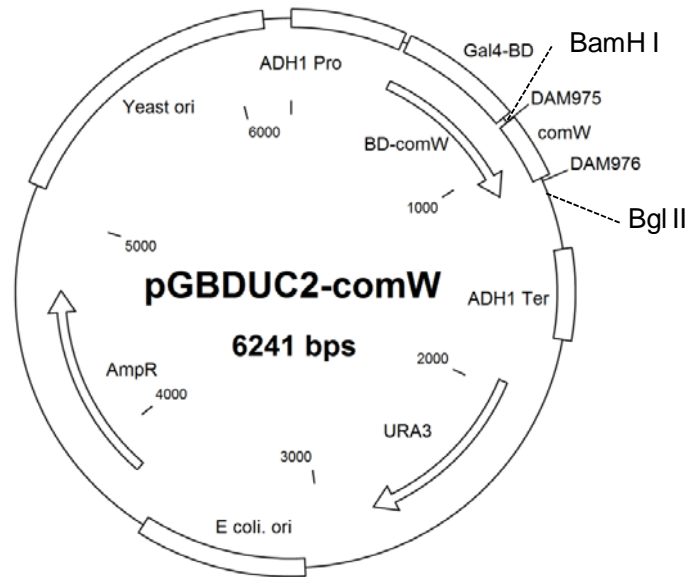
Appendix D (Continued)

Linker region of Gal4-BD and *comW*:

```
atcattgttt ccagtttctg tcaactgaca tagcggcctt aagggcccct aggaatacaa
>..............Gal4-BD............>>
s   s   n   k   g   q   r   q   l   t   v   s
>.............................BD-comW............................>
s   s   n   k   g   q   r   q   l   t   v   s   p   e   f   p   g   i   l   m
                                                            comW >>..>
                                                                    m

901 acaaaaaatt tatgagcaga tggctaattt ctatgatagt attgaagaag agtatggtcc
    tgtttttaa atactcgtct accgattaaa gatactatca taacttcttc tcataccagg
    >.............................BD-comW............................>
    l   q   k   i   y   e   q   m   a   n   f   y   d   s   i   e   e   e   y   g
    >.........................comW............................>
    l   q   k   i   y   e   q   m   a   n   f   y   d   s   i   e   e   e   y   g
```

Figure 48 Construction of chimeric plasmid pGBDUC2-*comW*. To insert *comW* into plasmid pGBDUC2, a copy of the *comW* was amplified with primer pair DAM975/976 and double-digested with BamHI and XhoI endonucleases. Meanwhile, circular pGBDUC2 plasmid DNA was also double-digested with BamHI and salI into linear vector. The insert and vector were then ligated together to make the chimeric plasmid, pGBDUC2-*comW*. Translation of the linker region showed that *comW* was inserted in the right ORF. The existence of *comW* in the chimeric plasmid was verified by enzyme digestion (BamHI and BglII), as shown in Fig. 20. The insert was cut off by BamHI, indicating that the ligation was accurate.
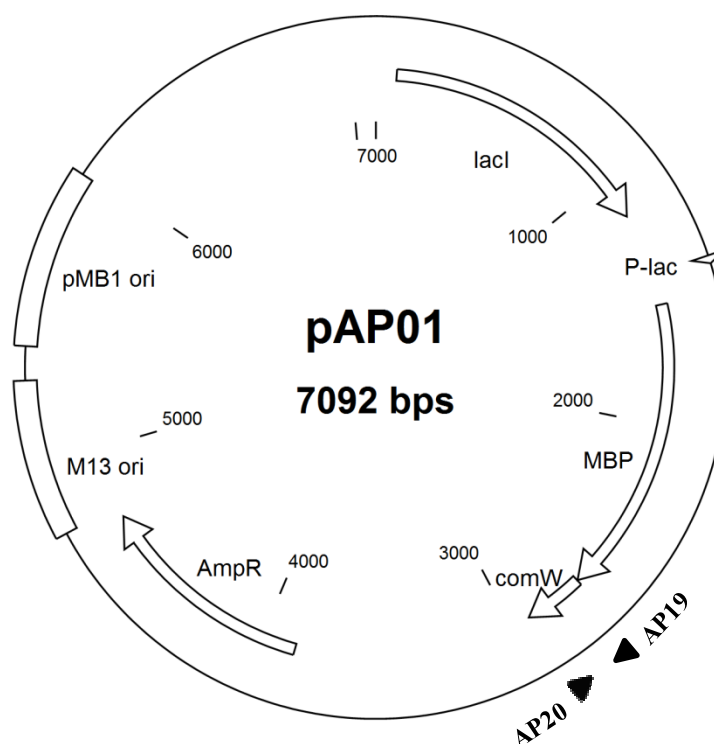
Appendix D (Continued)

Figure 49 Map of the plasmid pAP01 for purification of ComW from *E. coli*. The plasmid carries a fusion*, MBP-*comW*, which is induced by *lac* promoter with the addition of IPTG. The comW gene was amplified with primer AP19/20 and digested with BamHI endonuclease. The plasmid, pMAL-C2x, was digested by BamHI and XmnI to leave a sticky end and a blunt end. The insert and linear vector were then ligated to make pAP01. The structure of pAP01 was confirmed by enzyme digestion and sequencing.
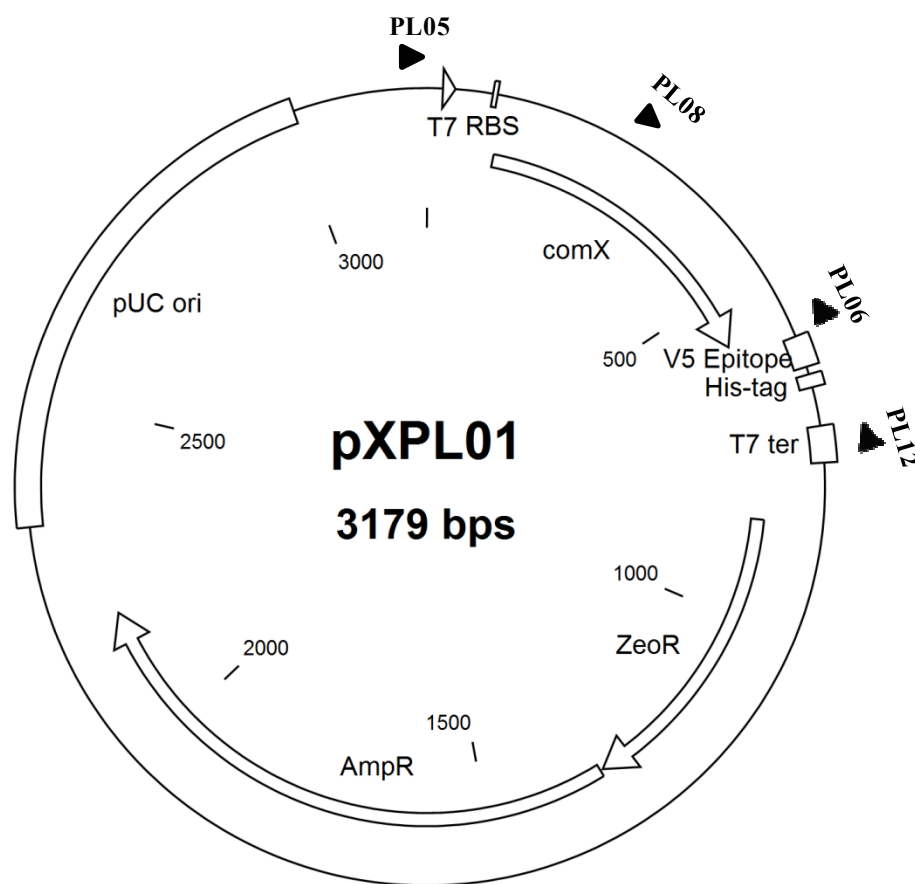
Appendix D (Continued)

Figure 50 Map of the plasmid pXPL01 for purification of ComX from *E. coli*. The plasmid carries *comX*-V5/His$_6$ cassette, whose expression is driven by T7 promoter and stopped by T7 terminator in E. coli. Two selective markers, ampicillin-resistant marker and zeocin-resistant markers, are included. pUC replication origin is used for its propagation in *E. coli*. The plasmid was made by amplifying a copy of *comX* gene with DAM284/MSL41 and inserting it into vector pCR-T7/CT-TOPO through blunt-end ligation. The structure of pXPL01 was confirmed by sequencing the PCR product (amplified with PL05 and PL12) with PL06 and PL08.
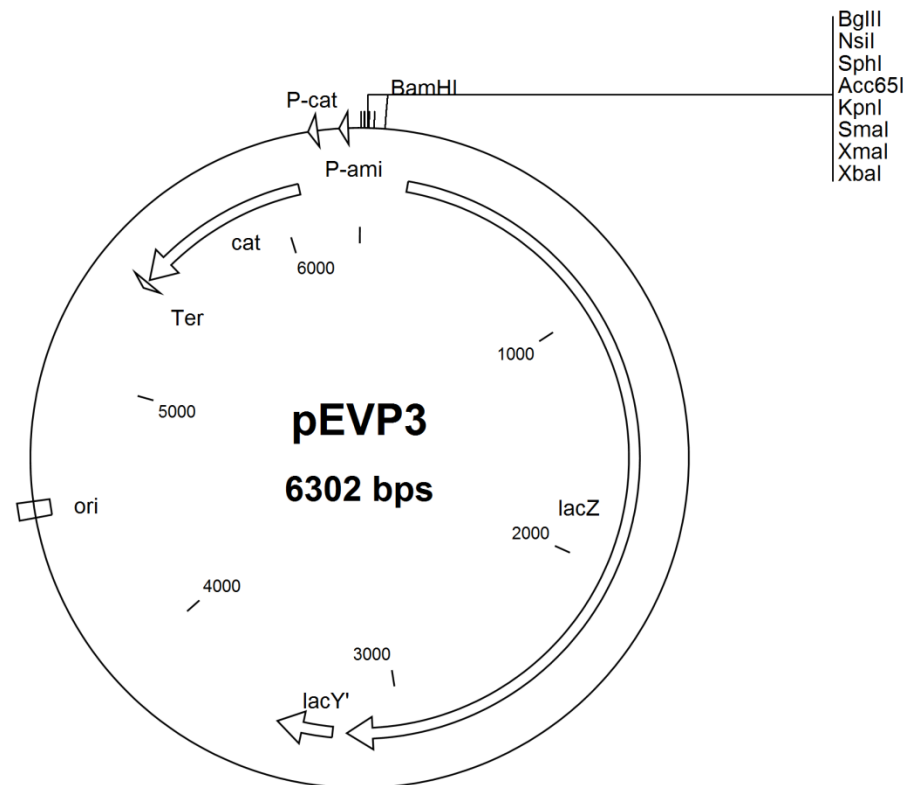
Appendix D (Continued)

Figure 51 Map of the plasmid pEVP3 for inserting *lacZ* reporter into pneumococcal genome. The plasmid carries a promoterless lacZ gene, the multiple cloning sites (BglII, NsiI, SphI, Acc65I, KpnI, SmaI, XbaI, and BamHI), and the pBR322 replication origin. The synthetic *cat* cassette contains chloramphenicol acetyltransferase gene, *cat* promoter (P-cat), *ami* promoter (P-ami) and the terminator (Ter).

VITA

NAME: Liming Weng

EDUCATION: Ph. D., Biological Sciences, University of Illinois, Chicago, IL, 2011

                   M.S., Pharmaceutical Analysis, Peking Union Medical College, Beijing, P. R. China, 2006

                   B.S., Pharmacy, Xi'an Jiaotong University, Xi'an, P. R. China, 1997

EXPERIENCE: Pharmacist, Department of Pharmacy, Renmin Hospital of Wuhan University, Wuhan, P. R. China, 1997-2003

                   Teaching Assistant, Department of Biological Sciences, University of Illinois at Chicago: Mendelian Genetics, Genetics Laboratory, Developmental Biology, University of Illinois at Chicago, 2006-2010

                   Research Assistant, Laboratory for Molecular Biology, Department of Biological Sciences, University of Illinois at Chicago, 2010-20011

MEMBERSHIP: The American Society for Microbiology

HONORS AND   Research Achievement Award, Department of Biological Sciences, university of
AWARDS: Illinois at Chicago, 2010.

                   Student Travel Award, 2nd International Conference on Gram-Positive Pathogens, Omaha, NE, 2010

                   Student Travel Award, 54th Wind River Conference on Prokaryotic Biology, Estes Park, CO, 2010

                   Student Travel Award, 5th International Conference on Gram-Positive Microorganisms, San Diego, CA, 2008

                   Awarded the University Scholarships in 1994, 1995 and 1996 respectively, Xi'an, China

PUBLICATIONS: **Liming Weng** and Donald Morrison. Exit from competence for genetic transformation in *Streptococcus pneumoniae* is regulated at multiple levels (manuscript in prep)

                   Liming Weng, Indranil Biswas, Donald Morrison. 2009. A self-deleting Cre-lox-ermAM cassette, Cheshire, for marker-less gene deletion in *Streptococcus pneumoniae*, J Microbiol Methods. 79(2009):353-357

                   Liming Weng, Yunshan Wu, et al. 2006. Identification on Chinese Traditional Medicine *scorpio* by powder X-ray diffraction fourier pattern method, Acta Academiae Medicinae Sinicae. 28(5):700-702

Wang shuchun,**Weng Liming**, et al. 2004. Identification on Chinese Traditional Medicine *hippocampus* by powder X-ray diffraction fourier pattern method, Northwest Pharmaceutical Journal, 19(6):248-250

Guo Xiaolin, and **Weng Liming**. 2002. Preparation and clinic observation of camptothecin gel, HubeiJournal of Chinese Traditional Medicine, 24(9):9-10

Guo Xiaolin, and **Weng Liming**. 2002. Preparation and observation of disintegrating tablets of ketoprofen, Chinese Journal of Hospital Pharmacy; 22(1):30-32

The Therapy of OTC Drugs, published by Hubei Science Publishing House, June 2002, ISBN 7-5352-2819-4, Wuhan (coeditor)

The Discipline of Practical Pharmacy Administration, published by Chinese Medicinal Science Publishing House, May 2000, ISBN 7-5067-2050-7, Beijing (coeditor)

ABSTRACTS:

Liming Weng and Donald A. Morrison (2009). "A Self-deleting Cre-lox Cassette for Streptococcus pneumoniae: Cheshire". 16th Annual Midwest Microbial Pathogenesis Conference. Purdue University. IN

Liming Weng, I. Biswas, and Donald A. Morrison (2009). "A Self-deleting Cre-*lox* Cassette for *Streptococcus pneumoniae:* CHESHIRE". 5th International Conference on Gram-positive Microorganisms. San Diego, CA

Liming Weng and Donald A. Morrison (2010). "DprA, a DNA processing protein, is also a ComX-dependent factor that shuts off competence in *Streptococcus pneumoniae*". 54th Wind River Conference on Prokaryotic Biology. Estes Park, CO

Liming Weng and Donald A. Morrison (2010). "DprA, a ComX-dependent repressor that shuts off competence in *Streptococcus pneumoniae*". International Conference On Gram-Positive Pathogens. Omaha, NE

Liming Weng and Donald A. Morrison (2011). "DprA, a ComX-dependent ssDNA binding and processing protein, is required for exit from competence in streptococcus pneumonia". 111th General Meeting of American Society for Microbiology. New Orleans, LA

Liming Weng and Donald A. Morrison (2011). "Interaction of ComX and ComW in development of competence for genetic transformation in S*treptococcus pneumoniae*". 55th Wind River Conference on Prokaryotic Biology. Estes Park, CO