Pneumococcal Transformation:
Mutations in the Primary Sigma Factor Bypass the Critical ComW Requirement

BY

YANINA TOVPEKO
B.S., Syracuse University, 2011

THESIS
Submitted as partial fulfillment of the requirements for the degree of Doctor of Philosophy in Biological Sciences (Molecular, Cell, and Developmental Biology) in the Graduate College of the University of Illinois at Chicago, 2016

Chicago, Illinois

Defense Committee:

David Stone, Chair
Donald Morrison, Advisor
Rachel Poretsky
Michael J. Federle, Center for Pharmaceutical Biotechnology
Nancy Freitag, Dept. of Microbiology & Immunology

This thesis is dedicated to the giants on whose shoulders I stand.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Dr. Donald A. Morrison for his support, encouragement, enthusiasm, and teaching. Dr. Morrison is the quintessential 'gentleman scientist', and his passion for science and education will forever leave a mark. He taught me that scientific research is an art, and that it can be performed with elegance and excellence. He has an amazing ability to identify the relevant information, to ask just the right questions, and to speak and encourage in such a way that makes everyone feel like they are really making accomplishments in science. I could not have asked for a better mentor, and I can only hope to spread his love for science to others.

I would like to thank the members of my thesis committee, Dr. Michael J. Federle, Dr. Nancy Freitag, Dr. Rachel Poretsky, and Dr. David Stone, for their encouragement, support, and mentorship. Dr. Stone is an amazing Committee Chair, and his suggestions were always helpful and thought provoking. I want to thank Dr. Poretsky for all her suggestions and help with writing. I want to thank Dr. Freitag for her thoughtful questions and for helping with acquiring my postdoctoral position. I want to thank Dr. Federle for adopting me as an honorary lab member during Dr. Morrison's adventures. I am truly honored to have such an amazing group of mentors.

I would like to thank current and former members of Dr. Morrison's and Dr. Federle's labs. Although I did not overlap with the former members of Dr. Morrison's lab, I would like to thank Ping Luo, Andrew Piotrowski, and Liming Weng for their wonderful theses and for the foundation on which my work is built. Nicole Inniss and Heidi Bai were wonderful labmates and always had helpful ideas during difficult experiments. I would like to thank our lab assistants, Anna Do, Om Bhetuwal, and Marlin Amy Halder, without them my experiments would never have succeeded. From the Federle lab, I want to thank Lauren Mashburn-Warren, Dr. Laura Case Cook,

ACKNOWLEDGEMENTS (continued)

Dr. Jennifer Chang, Dr. Tiara Perez-Morales, Reid Wilenking, Juan Cristobal Jimenez, Dr. Breah LaSarre, and Erin Shanker for all their suggestions, ideas, and always allowing me to share their equipment. There is no way I would have achieved this work without their help.

I would like to thank my friends in the department, especially Deanna Arsala, Lorna Cohen, Kristin Gallik, Emily Wyskiel, and Anna Do. Thank you for the support, suggestions, encouragement, your shoulder to cry on, your ears to complain to, your couch to laugh on, and your time. Without you, this program wouldn't nearly be so bright.

I would like to thank my friend Jennie Pikowsky, for her continual support and encouragement, both during college and afterwards. Her ability to be level-headed always reminded me to be steady and to continue to achieve whatever I wanted. I would like to thank Arvy Nowacki for his support and encouragement and his ability to help me see the bigger picture when experiments did not work no matter what I tried.

Lastly, I would like to thank my family, especially my mother and father. Their encouragement allowed me to pursue whatever I wanted and their support (both emotional and financial) made this degree possible. I could never begin to repay everything they have provided to me, but I hope that they know how thankful I am for them. Thanks, Mom and Dad!

YT

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (continued)

LIST OF ABBREVIATIONS

3D	Three-dimensional
α	alpha
A	adenine
a.a.	amino acid(s)
ABC	ATP-binding cassette
β	beta
β-gal	beta-galactosidase
bp	base pair(s)
BSA	bovine serum albumin
C	cytosine
CAT	casein hydrolysate yeast extract medium
CFU	colony forming unit
chi	chimera
cin-boxes	competence induced promoters
Cm	chloramphenicol
CSP	competence stimulating peptide
C-terminal	carboxy terminal
DNA	deoxyribonucleic acid
DNAseI	deoxyribonuclease I
dNTP	deoxynucleoside triphosphate
EDTA	ethylene diamine tetraacetic acid
Em	erythromycin

G	guanine
GP	glucose and phosphate
HGT	horizontal gene transfer
HRP	horseradish peroxidase
kb	kilobase
Kn	kanamycin
L	liter
min	minute
mu	molar unit
ng	nanogram
Nv	novobiocin
N-terminal	amino terminal
ω	omega
ORF	open reading frame
OD	optical density
PBS	phosphate buffer solution
PCR	polymerase chain reaction
PDB	protein data bank
PMSF	phenylmethyl sulfonyl fluoride
PVDF	polyvinylidene difluoride membrane
QS	quorum sensing
RNA	ribonucleic acid
RNAP	RNA polymerase

σ	sigma
SD	standard deviation
SDS	sodium dodecyl sulfate
Sm	streptomycin
SNP	single nucleotide polymorphism
Spc	spectinomycin
T	thymine
TBST	Tris-buffered saline-Tween
Tc	tetracycline
TCA	trichloroacetic acid
TE	Tris-EDTA
Tm	melting temperature
Tr	trimethoprim
WGS	whole-genome sequencing
WT	wild type

ABSTRACT

Streptococcus pneumoniae is an opportunistic pathogen that resides in the human nasopharynx and is naturally transformable, or able to take up and integrate exogenous DNA into its genome. Competence for genetic transformation is tightly regulated, transient, and occurs in two phases, early and late. The early genes encode a quorum-sensing system and peptide pheromone that signal an entire population to become competent in synchrony. One early gene encodes σ^X, the only known streptococcal alternative σ factor, which is responsible for coordinated synthesis of the late genes, those which are necessary for DNA uptake and recombination. In *S. pneumoniae*, elevated σ^X is insufficient for development of full competence without co-expression of a second competence-specific protein, ComW, which is regulated by the same pheromone circuit that controls σ^X. *comW* mutants display several phenotypes, a 10^4-fold reduction in the amount of transformants, a 10-fold reduction in σ^X activity, and a 10-fold reduction in the amount of σ^X protein. To identify proteins that may be interacting with ComW during competence, a suppressor screen was performed seeking mutants that were partially restored for transformation in the Δ*comW* mutant background. Whole genome sequencing of suppressor strains revealed ten different single-base substitutions in *rpoD*, the gene encoding the primary σ factor, σ^A, that each bypass the ComW requirement for transformation. Late gene expression was also restored in Δ*comW,* σ^A mutants. Eight of the ten single-base substitutions mapped to residues previously implicated in σ^A binding affinity to core RNA polymerase. Together, these data suggest that ComW increases σ^X access to core RNA polymerase, pointing to a role for ComW in σ factor exchange during genetic transformation in *S. pneumoniae*.

SUMMARY

Streptococcus pneumoniae is an opportunistic pathogen that resides in the human nasopharynx. It is naturally transformable, meaning it has the capability to take up DNA from the environment and integrate it into its genome. Competence for genetic transformation occurs during a brief, tightly regulated developmental phase that depends on two successive waves of gene expression; early and late. The early genes encode a peptide pheromone, an exporter, and a two-component signal transduction system, the coordinated expression of which leads to an entire population becoming competent in synchrony. The late genes encode the machinery necessary for DNA uptake and integration into the chromosome. The two waves of gene expression are linked by the only known alternative σ factor in the streptococci, σ^X. However, expression and accumulation of σ^X is not sufficient for the full competence response; it also requires expression of a second early gene, *comW*. ComW is important for the amount of transformants, the level of late gene expression, and the amount of σ^X during transformation. Despite various observations about the function of ComW, its specific role in regulating σ^X remains unclear.

To understand the critical role of ComW during transformation, I performed a suppressor screen in a *comW* mutant background to isolate rare individual clones that bypass the ComW requirement. This minimally biased approach provided a way to identify proteins involved with ComW during transformation. Once suppressors were isolated, whole-genome sequencing was used to identify the suppressor mutations, revealing ten different spontaneous mutations, each capable of increasing transformation efficiency in Δ*comW* strains. Every suppressor mutation mapped to *rpoD*, the gene encoding the primary σ factor, σ^A, causing ten amino acid substitutions. Eight of the ten substitutions mapped to sites in σ^A region 4 that are implicated in affinity for core

SUMMARY (continued)

RNA polymerase, suggesting that transformation efficiency is restored in σ^A mutants because of reduced σ^A affinity for core RNA polymerase.

Transformation efficiency is a complicated phenotype because it involves the concerted action of many proteins. In Δ*comW* σ^A suppressor strains, the yield of transformants was restored up to 10% of the WT level but the mechanism of restored transformation was uncertain. Because transformation depends on σ^X activity to initiate transcription of late genes, it is possible that σ^A suppressor mutants restore late gene expression and allows transformation to occur. To investigate whether the restoration of late gene expression was the source of the restored transformation efficiency, late gene expression was measured in σ^A suppressor mutant strains. Late gene expression was, in fact, restored in σ^A suppressor strains. Transformability, or the amount of transformable units per cell, was also restored in σ^A suppressor strains. In strains with the WT version of *comW*, we observed a decrease in the rate of late gene expression forty minutes after CSP induction. However, this rate decrease was not observed in any Δ*comW* mutant, regardless of whether it contained a mutant or WT version of σ^A. To determine whether the persistence in late gene expression resulted in a persistence of transformability, we measured competence kinetics in Δ*comW* σ^A suppressor strains. Competence kinetics, were normal in the σ^A suppressor mutants, except one case. The persistence of late gene expression suggests that ComW may also play a role in competence and late gene expression shut-off.

To efficiently create strains for genetic study, I used the high transformability of *S. pneumoniae* to edit its genome directly, without disrupting nearby genes. To facilitate this, the dependence of transformation efficiency on the amount and structure of donor DNA was investigated, for both genomic DNA donor amplicons and PCR amplicon donors. To compare the

dependence of transformation efficiency on the complexity of the donor, the transformation efficiency of PCR amplicons ranging from one-base substitutions to 1.2-kb cassettes was also monitored. Transformation efficiency increased with the increase in amount of donor DNA for both PCR and genomic donors, but the saturating amount varied by ~100-fold, indicating that transformation efficiency increases with the decrease in donor complexity. The near-100% transformation efficiency of PCR amplicons was used to perform markerless, scar-less, and selection-less editing on an essential gene. This successful, high-efficiency genome editing process was used to create most of the strains in this study.

ComW and σ^X are both required for competence, but much about ComW's role in transformation is still unknown. Data from this study suggests that ComW acts to facilitate competition of the alternative σ factor, σ^X, for access to core RNA polymerase. This points to a role for ComW as a novel σ factor activator, promoting σ factor exchange during transformation.

I. INTRODUCTION

1.1 Natural Transformation in Bacteria

1.1.1 Many bacteria are naturally transformable.

Natural transformation is one method of bacterial genetic exchange, in which bacteria take up DNA from the environment and integrate it into their genomes. This phenomenon was first observed by Frederick Griffith in 1928. Studying the bacterium *Streptococcus pneumoniae*, also known as pneumococcus or *Diplococcus pneumoniae*, he injected mice with either virulent heat-killed strains or live avirulent strains, and neither caused disease in the animals on their own. However, a mixture of the two strains did cause disease when injected subcutaneously into mice. Further study revealed that the bacteria isolated from the affected mice resembled the virulent strain. Griffith concluded that the live avirulent strain took up some element of the dead virulent strain, and he named this "the transforming principle" (Griffith 1928). The purification and characterization of the transforming principle of *S. pneumoniae* suggested that DNA is the hereditary material (Avery 1944), and was being exchanged between Griffith's rough and smooth strains.

Horizontal gene transfer (HGT) is one ancient mechanism of DNA exchange, with systems of HGT present in both Bacteria and Archaea (Polz et al. 2013). There are several mechanisms of DNA exchange, or horizontal gene transfer: conjugation, transduction, viral-like gene transfer agents, and transformation. Conjugation involves physical cell-to-cell contact by a projection between cells. Transduction is DNA transfer from one bacterium to another by bacteriophage. Gene transfer agents are virus-like particles that are be transferred from one organism to another. Natural transformation is the only DNA exchange mechanism that involves an organism's innate ability to take up DNA from the surrounding environment and incorporate it into its genome. When

bacteria are competent to take up DNA through natural transformation, they are said to be in the "competent state".

Initially, very few species of bacteria were reported to be competent for natural transformation (Lorenz & Wackernagel 1994). However, with the expansion and availability of genetic studies and genome sequences, it is now known that competence, or the ability to be naturally transformable, is widespread and is a shared feature of many diverse species across several phyla of bacteria. In fact, more than eighty species of bacteria have been shown to be naturally transformable (Johnston, Martin, et al. 2014), with surely more identified species to follow. With the widespread availability of whole-genome sequencing, full sets of genes required for competence have been observed in species that do not yet have any reported transformability. Interestingly, all sequenced species in the *Streptococcus* genus have the core set of genes necessary for transformation (Peterson et al. 2004; Fontaine et al. 2015; Johnston, Martin, et al. 2014), although not all have been shown to be transformable. This prompts the question as to whether these species retain these genes from a transformable ancestor and are no longer transformable, or whether they are transformable, but the necessary conditions for transformability in that species has not yet been identified. As each species occupies its own unique host and requires different conditions for growth, it is possible that transformation has not yet been observed only because the correct conditions have not yet been identified.

The length of time that bacteria are transformable varies. While transformation in the bacterium *Neisseria gonorrhoeae* can last indefinitely (Hamilton & Dillard 2006), most bacteria have a small window of transformability which is only induced under specific conditions. For example, *Bacillus subtilis* becomes competent during late log growth, and competence is induced by starvation (Jarmer et al. 2002). *S. pneumoniae* become competent during mid-log growth and

competence lasts for about 20 minutes after induction (Ween et al. 1999; Alloing et al. 1998; Tomasz & Hotchkiss 1964).

1.1.2 Why do bacteria transform?

There are several suggestions as to why bacteria have evolved to perform natural transformation (Redfield 1993). These suggestions include transformation as DNA repair mechanism (Charpentier et al. 2011), as a source of nutrients (Finkel & Kolter 2001), or as a source for genetic diversification (Ochman et al. 2000). Naturally transformable bacteria have been shown to induce competence under UV stress, a DNA-damaging agent (Charpentier et al. 2011). Repairing DNA damage could be the goal of natural transformation, but other studies have shown that survival is not improved in DNA-damaged cells that receive transforming DNA (Redfield 1993). Natural transformation may function to provide a mechanism for nutrient acquisition, as evidenced by *E. coli* mutants that experience fitness loss when unable to take up DNA, either heterologous or homologous to the recipient (Finkel & Kolter 2001). If the purpose of transformation is to take up DNA as a nutrient, it is puzzling that, in some bacteria such as *S. pneumoniae*, heterologous DNA is not imported or utilized. If transformation is solely to acquire nutrients, why would the source of the DNA matter? Also, up to 20% of bacterial genomes can contain foreign sequences indicating integration of foreign DNA (Ochman et al. 2000), suggesting that lateral gene transfer as a source of genetic diversification may be the utility of natural transformation.

For the streptococci, I propose the adaptive value of transformation is to sample DNA sequences from nearby relatives to improve genetic diversity. This idea is supported by the presence of a highly efficient recombination system requiring integrating DNA to have homology to the recipient chromosome, indicating that transformation relies on having related species in the

nearby environment. Also, the argument for genetic transformation as a gene sampling mechanism is supported by the presence of a fratricidal protein that is produced during competence that lyses only non-competent pneumococci, suggesting that a DNA releasing mechanism exists to produce genomic DNA from nearby related bacteria (Eldholm et al. 2009). Regardless of why bacteria have evolved to exchange DNA, transformation is important to study because in many organisms, DNA exchange leads to drug-resistance and immune escape (CDC 2013; Swartley et al. 1997; Wyres et al. 2013). For example, in the streptococci, recombination events at the capsule locus lead to vaccine escape and evasion of the host immune response (Golubchik et al. 2012; Brueggemann et al. 2007; Jefferies et al. 2004).

1.2 Transformation in the streptococci

1.2.1 The streptococci

Streptococci are facultative anaerobes, they are Gram-positive, alpha-hemolytic, oval shaped bacteria that usually exist in two cells bound at the ends, in either short or long chains depending on the species and growth conditions. Species in the streptococcus genus occupy a wide variety of hosts such as fish, pigs, cows, and humans, as well as many other environments such as milk and cheese (Facklam 2002; Martín et al. 2015; Tsakalidou et al. n.d.). The streptococcus genus has been reclassified and reorganized many times. Rebecca C. Lancefield sorted streptococci into groups based on their hemolytic qualities and surface antigens in 1933, and since then, her studies and those of others' refined this sorting into more discrete groups using newly available technology (Lancefield 1933; Lancefield 1962; Facklam 2002). In 1984, the streptococci were further separated into three genera, Streptococci, Lactococci, and Enterococci (Schleifer & Kilpper-Balz 1984). With many changes in the taxonomy, nomenclature, and organization of this genus, a guide was created suggesting the use of many tools to type streptococcal species, allowing

for more careful identification of new isolates (Facklam 2002). Streptococci have also been sorted into species based on 16S RNA sequencing (Kawamura et al. 1995). More recently, streptococci are sored based on whole genome sequencing, which is currently the most accurate way to delineate a phylogeny of species (Richards et al. 2014) (Fig. 1).

1.2.2 Transformation in streptococci

In the streptococci, the mechanism of transformation involves attack on other streptococci, DNA binding, uptake, protection, and recombination (Fig. 2). In *S. pneumoniae*, double-stranded donor DNA binds at the at the midcell (Bergé et al. 2013), where the EndA nuclease degrades one strand and the remaining single strand is brought into the cell (Bergé et al. 2013; Puyet et al. 1990; Mejean & Claverys 1993). The single strand is coated with single-strand binding protein B (SsbB), which prevents degradation (Morrison & Mannarelli 1979; Morrison et al. 2007). This complex is brought to the chromosome and recombined based on sequence homology at the chromosomal locus. In most transformation events, a single strand is replaced (Gurney & Fox 1968; Guerrini & Fox 1968). Once cell division occurs, one transformant daughter cell inherits new sequences that were recombined into the chromosome and WT daughter cells retain the endogenous copy of the sequence, if only one strand of DNA was replaced at one locus (Porter & Guild 1969). However, if more than one copy of a region of DNA is replaced, on both strands, the frequency of transformed daughter cells in the population will be higher than for a single copy. (Fig. 3) (Porter & Guild 1969; Louarn & Sicard 1968; Tiraby et al. 1975).

This transformation process requires the coordinated expression of many genes and proteins required for the above mentioned transformation steps. Their concerted expression relies on the only known streptococcal alternative sigma factor, σ^X, which transcribes the genes necessary for the competent state, or X-state (Claverys et al. 2006). Every streptococcus species

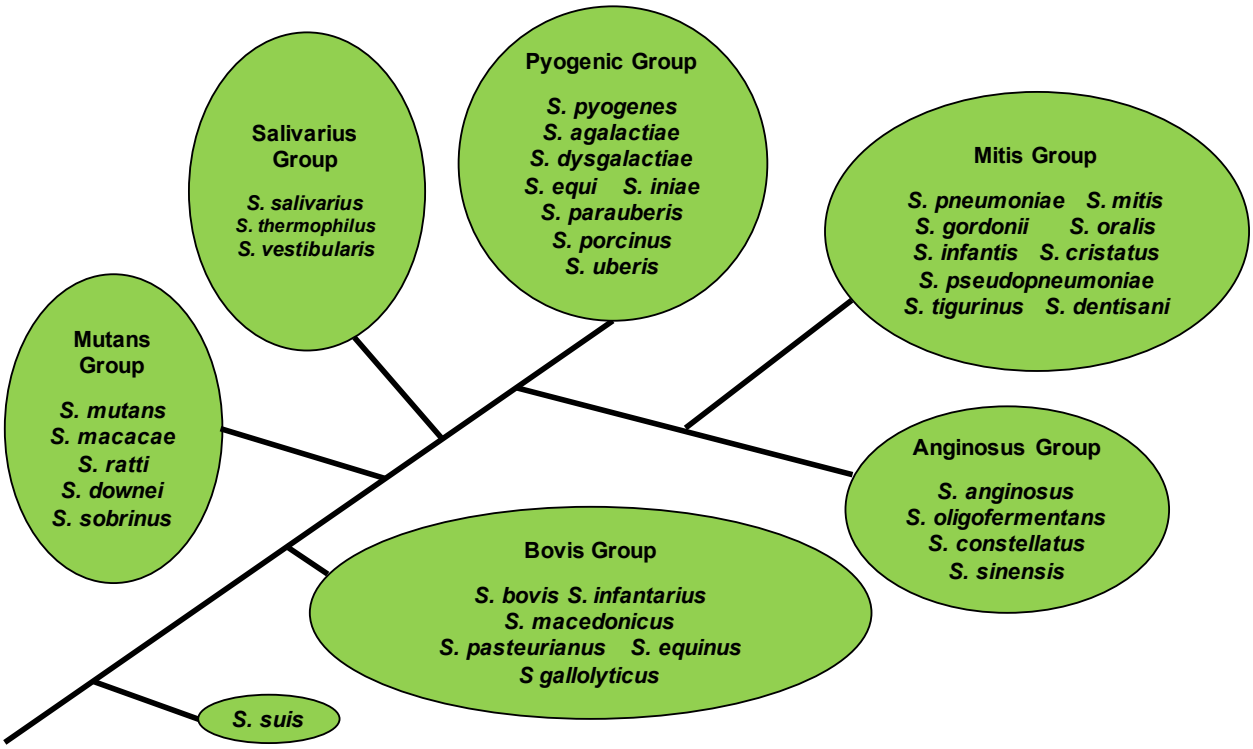


Figure 1. Simplified phylogeny and grouping of streptococcal species. The streptococci are divided into several groups based on relatedness and phylogeny is also assessed using sequencing. This phylogeny is adapted from several sources using 16S RNA sequencing (Kawamura et al. 1995) and whole genome sequencing (Richards et al. 2014). Green ovals, groupings; black vertical text, group name; black italic text, species in the groups; black lines, phylogenetic branches.

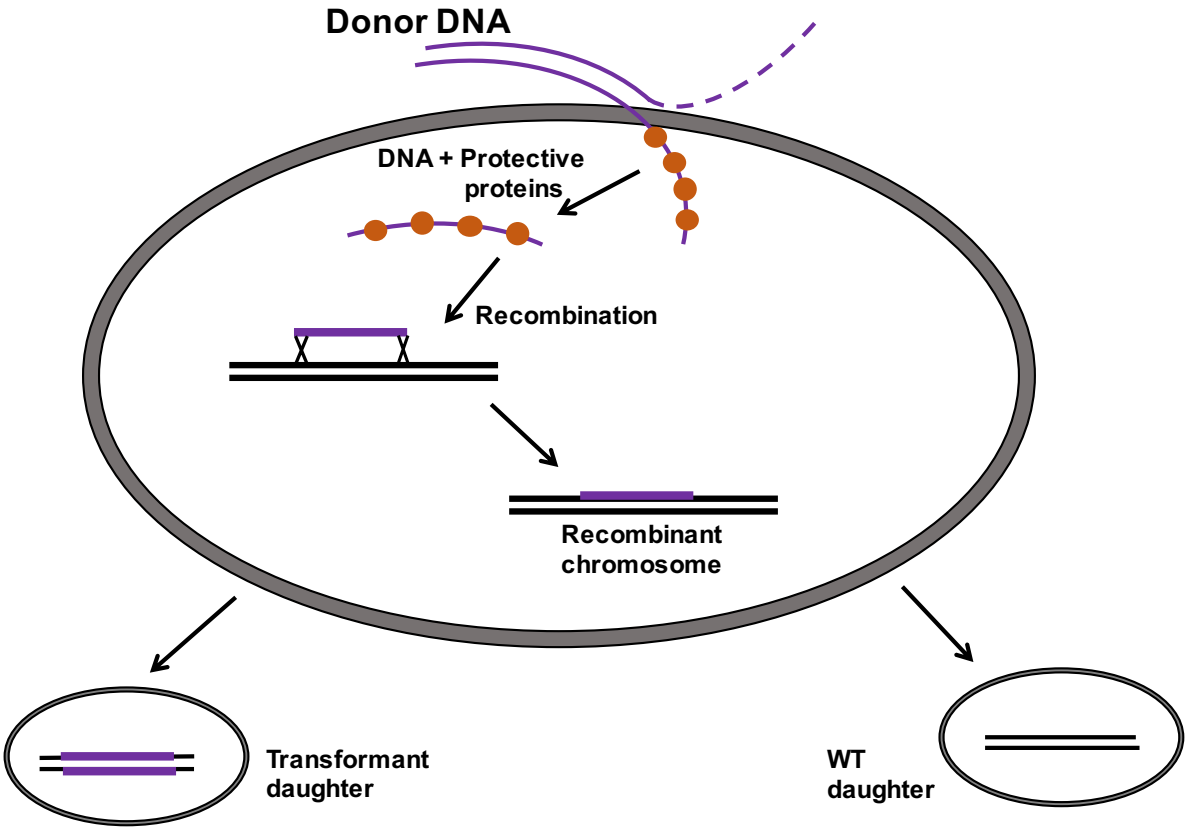


Figure 2. General transformation scheme in streptococci. Transformation requires DNA uptake, protective proteins, and recombination at the chromosome. A donor DNA duplex binds at the midcell, where one strand is degraded and the other is brought into the cell. The single strand is coated with proteins to protect from degradation, and the DNA is brought to the chromosome and recombined based on homology of sequences. Upon division, there is one transformant daughter cell with new sequences that recombined into the chromosome and one WT daughter cell that retained the endogenous copy of the sequence. Gray oval, streptococcal cells; purple, donor DNA; orange circles, DNA binding proteins; black horizontal lines, chromosomal DNA; black arrows, steps in transformation.

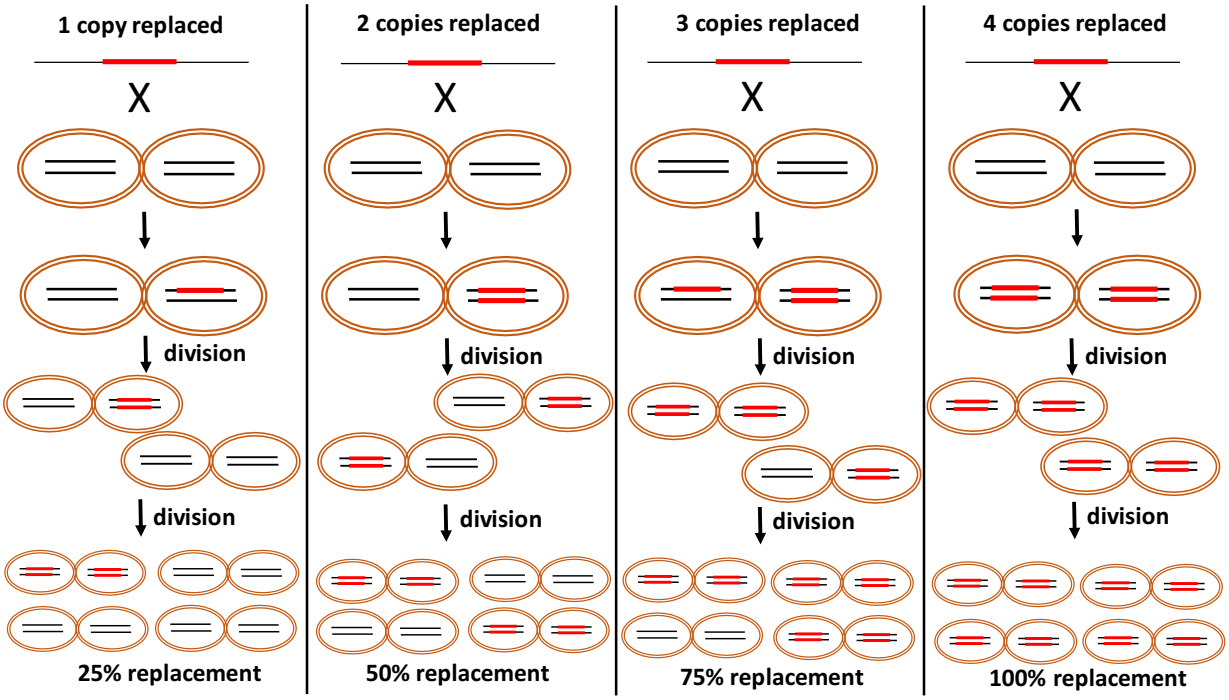


Figure 3. Examples of the number of loci that can be replaced during transformation and the resulting percentage of replacement in the population. *S. pneumoniae* exists as diplococci, and four copies of a locus can be present in a single colony forming unit (CFU) at any one time. If one loci or copy is replaced by donor DNA during transformation, 50% of cells will have the recombined loci after the first division. However, after the second division and segregation of DNA, 25% of the cells will have the transformed loci. If two copies are replaced, 100% of cells will have replaced loci after the first division, but a second division reveals 50% will have replaced loci. For three copies replaced, the final replacement after two divisions is 75%. For four copies, the final replacement is 100% of cells. Black lines, DNA; red boxes, donor DNA; red boxes inside cells, replaced loci; orange circles, streptococci.

examined to date has been found to have at least one copy of *comX* (Martin et al. 2006), the gene that encodes σ^X, and interestingly, most streptococci have two copies of the *comX* gene (Lee & Morrison 1999; Martin et al. 2006). While competence was, at first, only reported in *S. pneumoniae*, many streptococcus species have now been shown to be transformable (Johnsborg et al. 2007), including *S. mutans* (Li et al. 2001; Mashburn-Warren et al. 2010), *S. thermophilus* and *S. salivarius* (Fontaine et al. 2010), and *S. infantarius* and *S. macedonicus* (Morrison et al. 2013). Not all streptococci have been reported to be transformable in laboratory cultures, but it is possible that in these species transformation requires conditions that more closely mimic the natural habitat of the bacterium. For example, a recent study reported that *S. pyogenes* is transformable in biofilms grown on human epithelial cells (Marks et al. 2014). More studies of transformation in streptococci may reveal that most or all the streptococcus species are transformable, if the necessary conditions are uncovered (Håvarstein 2010).

1.3 Competence induction in *S. pneumoniae* requires quorum sensing and a two-component signal transduction system.

Bacteria exist in a gamut of population sizes, ranging from single to a few cells (low population density) to larger communities (high population density). Although they are unicellular organisms, they respond to their environment as whole populations. They do this by employing secreted signals, generally termed 'autoinducers' (Bassler & Miller 2013). Different bacteria use different molecules as autoinducers. Autoinducers can be short peptides, N-acyl homoserine lactones, or boron-containing autoinducers, and they differ between Gram-negative and Gram-positive bacteria (reviewed in (Miller & Bassler 2003)). The production, secretion, detection, and response to these signals is called quorum sensing (QS). At low population densities, these signals are also at low concentrations in the surrounding environment. However, as population density

increases, the peptide molecule also increases in concentration (Fig. 4). Autoinducers allow for cell-to-cell communication, and for the population to respond as a whole rather than as individual cells (Bassler & Miller 2013).

QS mechanisms that regulate competence in the streptococci fall into two groups, ComRS and ComCDE (Mirouze et al. 2013; Johnston, Martin, et al. 2014). The anginosus and mitis groups use ComCDE, which involves simultaneous export and cleavage of a small peptide that is sensed externally (discussed below). The mutans, salivarius, pyogenic, and bovis groups use ComRS, which involves peptide processing and excretion, then importation by an oligopeptide permease, reviewed in (Fontaine et al. 2013). Both mechanisms of QS control the expression of the alternative σ factor, σ^X (Johnston, Martin, et al. 2014; Fontaine et al. 2010; Fontaine et al. 2013; Mashburn-Warren et al. 2010; Morrison et al. 2013).

In *S. pneumoniae,* the QS mechanism for transformation is encoded by two genetic loci, *comAB* (Chandler & Morrison 1988; Hui & Morrison 1991; Hui et al. 1995) and *comCDE* (Pestova et al. 1996; Håvarstein et al. 1996; Martin et al. 2013; Martin et al. 2010). These operons are transcribed at a basal level during growth from weak promoters (Martin et al. 2013). ComA is an ABC transporter protein (ATP-binding cassette), an ATP-dependent transporter protein (Hui & Morrison 1991) and ComB is the catalytic subunit (Hui et al. 1995). Together, they export and cleave the pro-peptide translated product of *comC*, a 43 aa (amino acid) protein. This peptide is cleaved at a double-glycine motif, and the mature 17 aa secreted peptide is called CSP (competence stimulating peptide) because it induces competence in *S. pneumoniae* (Havarstein et al. 1995) (Fig. 5). CSP exists in two allelic variants, CSP-1 and CSP-2, distinguishable by eight conservative aa substitutions. They exist roughly in equal proportion amongst pneumococcal strains, and strains typically respond only to their own CSP variant (Pozzi et al. 1996).

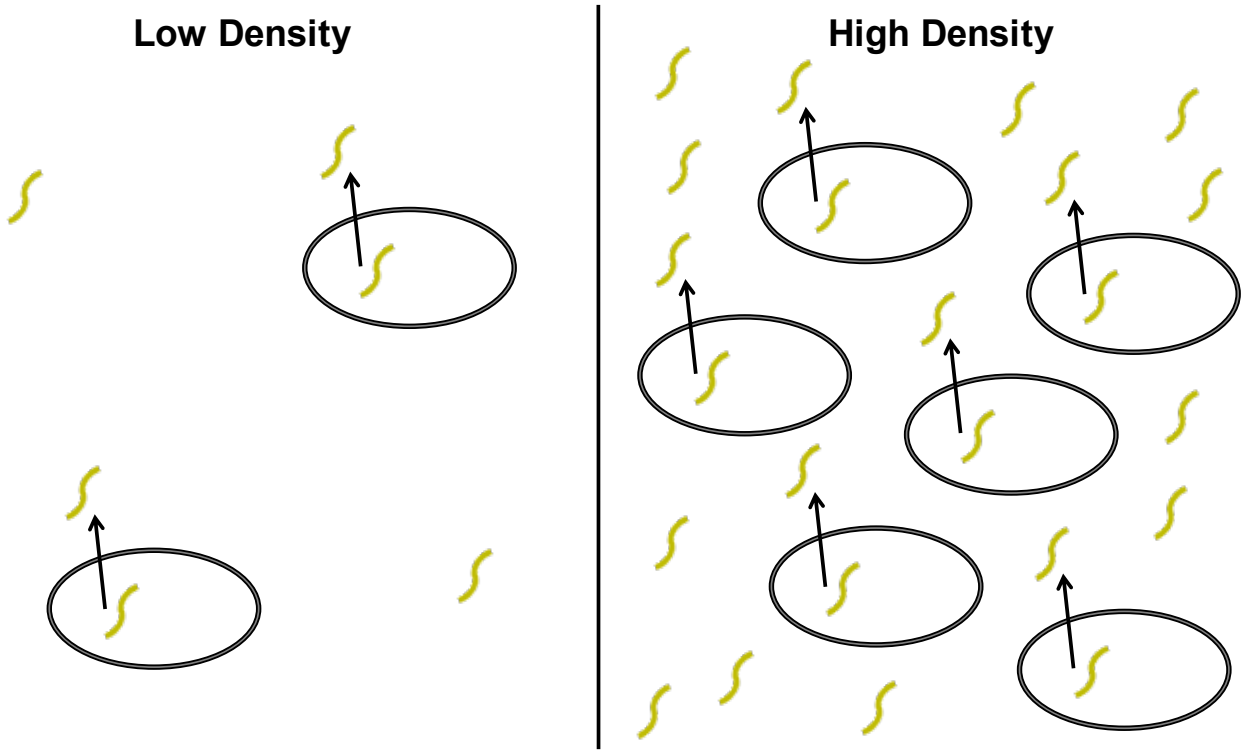


Figure 4. Autoinducer molecule increase with bacterial population density increase. At low bacterial population densities, any inducer molecules that are secreted are present at low concentration. Increases in population density increases the concentration of secreted molecules, allowing for synchronous population response. Gray ovals, bacterial cells; yellow curved line, secreted autoinducer; black arrow, secretion from inside of cells to environment.

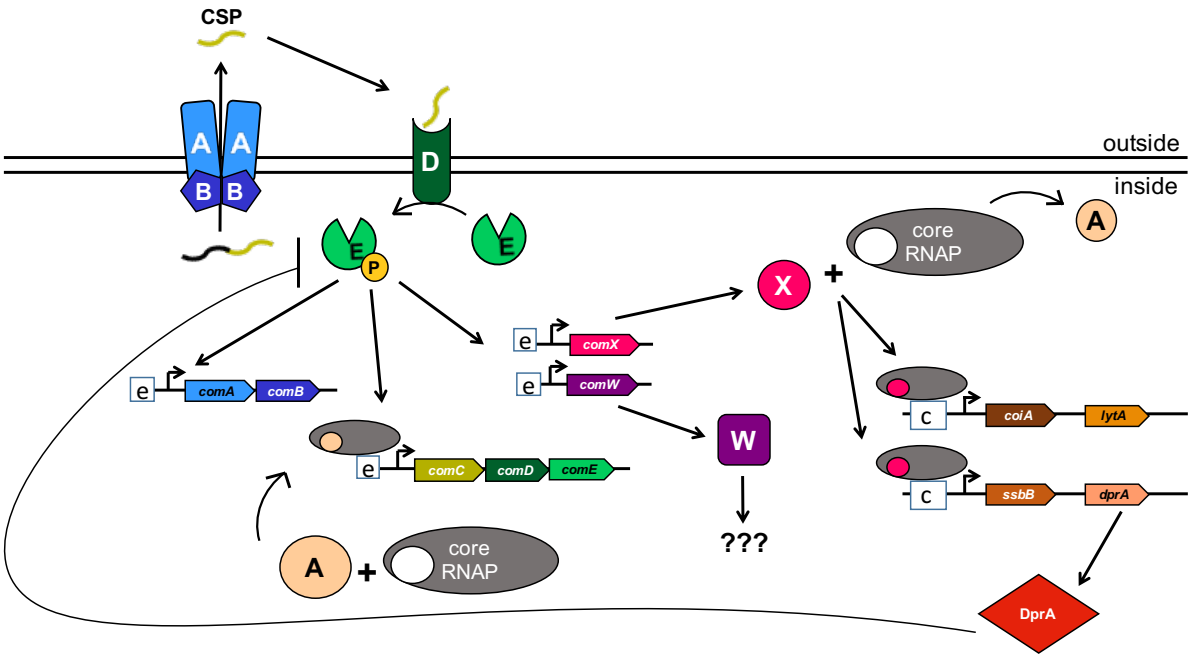


Figure 5. Induction of early genes and late genes by σ^X in *S. pneumoniae*. The competence genes *comAB* and *comCDE* are are transcribed at a basal level from a poor promoter by RNA polymerase bound to the primary σ factor, σ^A. The pro-peptide product of *comC* is exported out of the cell by ComAB, the 17 amino acid mature product is CSP (competence-stimulating peptide). ComD, a histidine kinase receptor, senses CSP and phosphorylates its cognate response regulator, ComE. Phosphorylated ComE binds at the ComE-box to upregulate transcription and induce a positive feedback loop. Two genes that have this same induction pattern are *comX* and *comW*. σ^X binds to core RNA polymerase to recognize and transcribe from the non-canonical promoter, or cinbox, upstream of late competence genes. ComW is a small protein essential for transformation. DprA is a late protein that is a negative shut-off regulator of ComE. Inside, inside of cell; outside, outside of cell; blue rectangles, ComA; small dark blue pentagons; ComB; green cylinder, ComD; light green circles, ComE; yellow circle, phosphate; gray oval, core RNAP, beige circle, σ^A; magenta circle, σ^X; purple square, ComW; brown, light brown, orange, peach, late genes; red, DprA; elongated pentagons, genes; bent arrow, promoter; curved or straight arrow, location of action; boxed e, ComE-box; boxed c, cinbox; double black line, cell wall/membrane.

CSP is sensed by a histidine kinase receptor ComD, and upon sensing CSP, ComD phosphorylates its cognate response regulator, ComE (Håvarstein et al. 1996; Martin et al. 2010; Martin et al. 2013). The phosphorylated ComE (ComE-P) binds upstream of promoters at a direct repeat centered at the -40 area, also called the ComE-box (Peterson et al. 2004; Ween et al. 1999). *comAB*, *comCDE,* and 6 other operons which contain the ComE-box upstream of their promoters are called the early genes because they are the first set of genes turned on during response to CSP. The early competence genes are normally transcribed at a basal level allowing for a small amount of ComD and ComE to be present to sense CSP (Martin et al. 2010). Once CSP is sensed and ComE-P binds upstream of *comAB* and *comCDE*, it incites a positive auto-feedback loop, allowing for upregulation of these genes, which are transcribed at high levels only during competence (Martin et al. 2013; Martin et al. 2010; Oggioni & Morrison 2008). This concerted induction and upregulation of the early genes prompts the swift increase in the level of CSP and synchronous competence in an entire culture or population of cells.

1.4 The alternative sigma factor, σ^X, is the link between early and late gene expression.

One early gene of paramount importance for transformation is *comX*. *comX* encodes the only known alternative sigma factor in the streptococci, σ^X (Lee & Morrison 1999). Every species of the *Streptococcus* genus that has been investigated so far has σ^X, and in every transformable species it is also linked to the expression of late competence genes (Johnston, Martin, et al. 2014). *S. pneumoniae* belongs to the mitis group. The mitis and anginosus groups (Fig. 1) have orthologous ComABCDE and CSP QS systems as the upstream regulators of competence (Håvarstein 2010). In the other groups of streptococci, *sigX*, the ortholog of *comX* in *S. pneumoniae*, is regulated by a different QS peptide system, using proteins ComRS, which encode a peptide pheromone and intracellular receptor (Håvarstein 2010). Precise regulation of σ^X in any

species is important because some of the genes it encodes are involved in fratricide and cognate immunity to these fratricidal proteins (Eldholm et al. 2009).

σ^X is the link between early and late gene expression during competence. There are two copies of *comX* in *S. pneumoniae*, and either copy is sufficient for competence induction (Lee & Morrison 1999; Luo & Morrison 2003; Luo et al. 2003). σ^X accumulates to a high intracellular level in response to CSP induction (Luo & Morrison 2003; Luo et al. 2003; Lee & Morrison 1999). σ^X binds to and forms a holoenzyme with core RNA polymerase (RNAP), allowing for the recognition and transcription from a noncanonical promoter sequence called the ComX-box or cinbox. This is found upstream of at least 23 genes, such as those necessary for DNA uptake and recombination, called the late competence genes (Lee & Morrison 1999; Luo et al. 2003; Luo & Morrison 2003; Morrison 2007; Peterson et al. 2004; Campbell et al. 1998) (Fig. 5). σ^X is responsible for transcribing all of the late genes necessary for transformation, and forms a holoenzyme with RNAP when the primary σ factor, σ^A, is not bound. The concerted expression of the early genes, σ^X, and late genes results in a peak of transformants at 20 minutes after CSP induction, and a shut-off of transformation is visible at 40 minutes (Fig. 6). One of the late genes transcribed by σ^X, *dprA*, encodes the negative shutoff regulator DprA, which binds to ComE to induce the shutoff of competence (Fig. 5) (Martin et al. 2013; Mirouze et al. 2013).

1.5 σ (sigma) factors

1.5.1 The primary σ factor

During bacterial transcription, the σ factor is a protein responsible for recognizing the -10 and -35 elements of the promoter and transcription initiation (Paget & Helmann 2003). σ was first identified as a DNA-binding transcription initiation factor in *Escherichia coli*, and termed σ^{70} because of its size under separation (Burgess et al. 1969). However, in mostly all other bacteria,

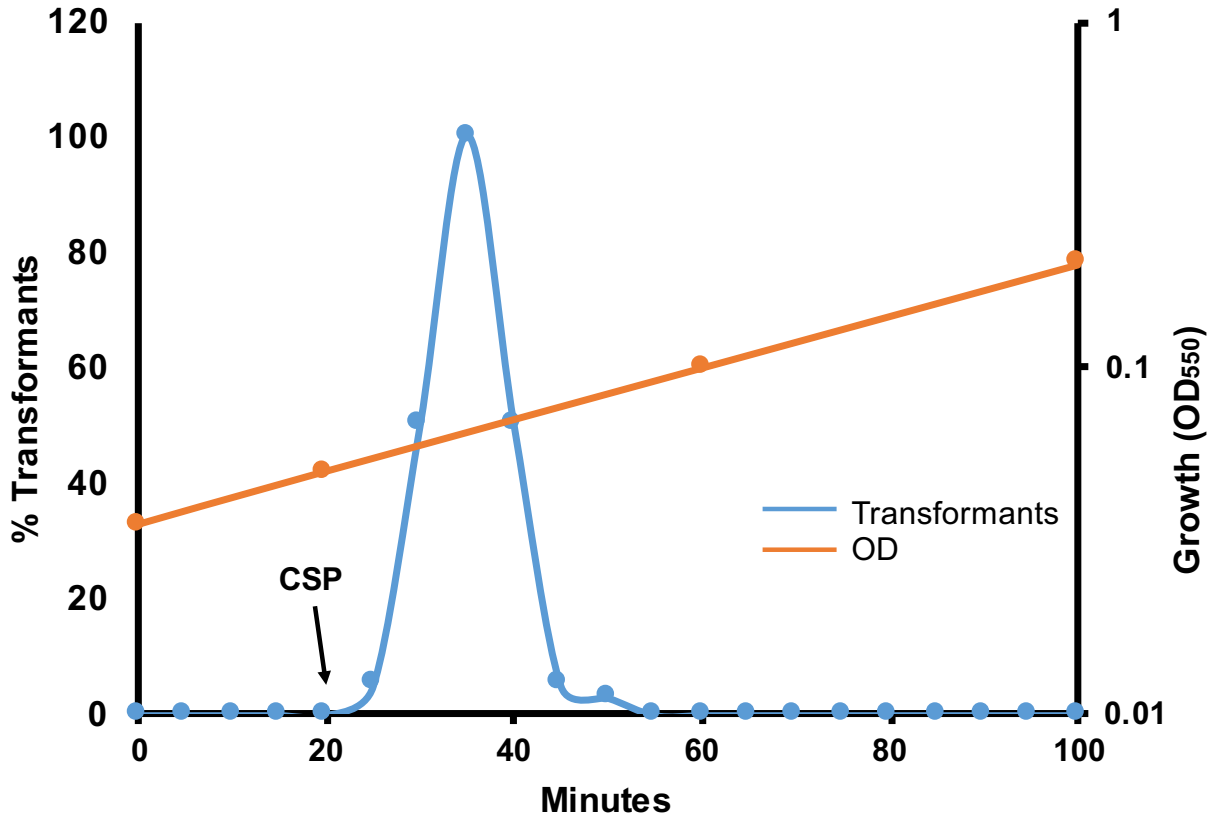


Figure 6. Kinetics of transformation during growth in *S. pneumoniae*. After CSP induction, the amount of transformants from a competent culture will rise and peak at 20 minutes after induction, and no transformability is visible after 40 minutes post-induction. Percent transformants is measured as the number of colonies on a selection plate divided by the number of cfu at each time point. Growth is measured by absorbance (OD) at 550 nm. Blue line, amount of transformants after CSP induction; orange line, OD_{550}; black arrow, point of CSP addition to a log phase culture.

the primary σ factor is called σ^A. σ^A is the master regulator of global transcription in bacteria. σ^A binds to core RNA polymerase, which consists of two α subunit, one β subunit, one β' subunit, and one ω subunit (Murakami & Darst 2003). Together, the transcription initiation factor σ and core RNAP bind DNA (Murakami, Masuda & Darst 2002) and perform transcription of genes from DNA to RNA (Murakami, Masuda, Campbell, et al. 2002).

The primary σ factor is divided into four domains, or regions, based on different functions and roles for each domain (Vassylyev et al. 2002). Region 1 lies at the N-terminus of σ and is responsible for transcription initiation and promoter recognition (Wilson & Dombroski 1997). Region 2 is responsible for recognizing and binding the -10 element of the prokaryotic promoter (Barne et al. 1997). Region 3 is responsible for binding to the β subunit of core RNAP (Severinov et al. 1994). Region 4 is a hotbed of activity and is responsible for affinity for core RNAP (Nickels et al. 2005), binding regulatory proteins (Lonetto et al. 1998; Dove et al. 2003), and binding the -35 element of the prokaryotic promoter (Lonetto et al. 1992).

The primary σ factor is responsible for recognizing the canonical promoter and transcribing almost all genes (Reznikoff et al. 1985). It has the highest affinity for core RNAP when compared to alternative σ factors (Maeda 2000). Alternative σ factors bound to core RNAP recognize and transcribe from promoters that range from slightly different from the canonical promoter to non-canonical promoter, and have lower affinities for core RNAP when compared to the primary σ (Österberg et al. 2011; Maeda 2000).

1.5.2 Alternative σ factors

Alternative σ factors bind core RNAP and allow for promoter discrimination, resulting in transcription from promoters that are different from the canonical promoter recognized by the primary σ factor. Some species of bacteria, such as *S. pneumoniae* only have one alternative σ,

while other species have over seventeen identified alternative σ factors (Britton et al. 2002), and other bacteria such as *Streptomyces coelicolor* have over sixty identified alternative σ factors (Gruber & Gross 2003). These σ factors have different roles during the life cycles of the bacterium, for example: σ^X is responsible for transcribing the competence specific genes in *S. pneumoniae* (Lee & Morrison 1999), σ^S is responsible for transcribing genes necessary during stationary phase in *E. coli* (Zhou & Gottesman 1998), σ^{30} or AlgU is responsible for transcribing alginate in *Pseudomonas aeruginosa*, allowing for mucoidy in cystic fibrosis patients (Deretic et al. 1994), and the many σ factors such as σ^E and σ^K necessary for sporulation in *B. subtilis* (Haldenwang 1995). Alternative σ factors must compete with σ^A for binding to core RNAP and mostly have lower affinities for core RNAP compared to the housekeeping σ (Farewell et al. 1998; Maeda 2000).

Because alternative σ factors control specific responses that often involve a huge expense of energy for the bacterium, their availability and function is tightly regulated. First, the availability of the alternative σ factors is usually under transcriptional control, requiring environmental signals to induce transcription of the alterative σ by the primary σ (Österberg et al. 2011). Then, even when a σ factor is present, it is acted upon by regulatory proteins either to enhance or prevent its activity. Some alternative σ factors are kept in an inactive state by sequestration, such as the protein RsbW that sequesters σ^B in *Staphylococcus aureus* (Miyazaki et al. 1999). Once a σ factor is present, many are subject to regulation by proteases such as σ^X degradation by ClpEP in *S. pneumoniae* (Sung & Morrison 2005). Some alternative σ factors require a small accessory protein to bind before they bind core RNAP, such as Crl binding σ^S in *E. coli* (Banta et al. 2013). There are many different mechanisms of coordinating alternative σ factor activity, and understanding the

specific regulation of any σ factor sheds light on the genetic pathways under the control of that σ factor.

1.5.3 Regulation of σ factor activity can be studied using suppressor screens.

A suppressor screen is a biological technique that searches for mutations that alleviate a known mutant phenotype. The identity of the suppressor often reveals key information about the role of the protein in question. A review of suppressor studies involving σ factors showed two classes of suppressors that occur relating to alternative σ factor regulation.

The first class contains mutations in the primary σ factor that reveal reduced regulation of several pathways, leading to increased transcription from alternative σ factors (Table I). For example, in the absence of the alarmone ppGpp, *E. coli* cells are lacking in their nutrient limitation response, and this was traced to the silencing of σ^N activity. A study by Laurie et al. revealed that mutations in σ^A compensate for and overcome the loss of ppGpp, and these σ^A mutations result in increased transcription from σ^N without the need for ppGpp (Laurie et al. 2003). Other studies exploring regulatory pathways also revealed increased alternative σ factor activity when there were suppressor mutations in the primary σ factor (Yin et al. 2013; Nickels et al. 2005). The implication from such results is that mutations resulting in the decrease of σ^A activity result in increases in alternative σ factor activity and overcome the need for regulatory proteins.

The second class of suppressors are found in regulatory proteins, revealing proteins that controlling alternative σ factor activity (Table I). For example, in *B. subtilis*, there is decreased activity of the alternative σ factor, σ^B, in RbsX mutants. It was unknown how RbsX regulated σ^B activity, but suppressor mutations in *rbsU* revealed that RbsU was necessary for RbsX regulation of σ^B (Voelker et al. 1995). A similar study implicated RssB in PhoP-dependent accumulation of σ^S in *E. coli*, where *rssB* suppressor mutations restored the reduced accumulation of σ^S in *phoP* mutants

TABLE I. Studies and identified suppressors of σ factors that shed light on regulation of σ factors.

Goal of Study[a]	Identified Suppressor[b]	Suppressor phenotype[c]	Implication[d]	Organism[e]	Source
Understand the temperature sensitive phenotype of CpbA/DnaJ mutants	Two copies of rpoD (σ^A)	Loss of temperature sensitivity	Increasing the amount of σ^A and transcription from σ^A rescues reduced cellular process in temp. sensitive mutants	*E. coli*	Shiozawa, 1996
Investigate the silencing of σ^N caused by the loss of ppGpp	σ^{70}-ΔDSA(536-538)	Increased σ^N activity, reduced σ^A activity	Reduced σ^A affinity for core allows for increased transcription from alternative σ factors	*E. coli*	Laurie, 2003
Understand the upregulation of transcription activation in λCI mutants	σ^{70}-R596	Reduced transcription from λ promoters	Mutating a residue in σ^A destabilizes promoter recognition and binding to core RNAP	*E. coli*	Dove, 2003
Residues important for binding to core RNAP	σ^{70}-L607P	Reduced σ^{70} interaction with the β-flap of core RNAP	This residue is important for σ affinity for core RNAP	*E. coli*	Nickels, 2005
Residues important for binding ssDNA and promoter melting	σ^{70}-Y184A, Y189A, Y193A	Mutated σ has reduced promoter melting activity	These residues bind the promoter and are important for promoter melting	*E. coli*	Huang, 1997
Mutations that reduce susceptibility to antibiotics	σ^{70}-G336C	Increased σ^B activity, reduced susceptibility to vancomycin	Increase of stress response gene transcription from σ^B, reduced activity from σ^{70}	*B. subtilis*	Lee, 2014
Investigating the mechanisms of increased alginate production by *P. aeruginosa*	Decreased σ^A expression	Increased expression from AlgU, the alt. σ responsible for alginate production	Reduced σ^A activity allows for increased transcription from alternative σ factors	*P. aeruginosa*	Yin, 2013
To identify decreased σ^B function in RbsX mutants	Insertion mutation in rbsU	Restored activity of σ^B in rbsX, rbsU mutants	RbsU is necessary for RbsX regulation of σ^B activity	*B. subtilis*	Voelker, 1995
To investigate the loss of PhoP-dependent accumulation of σ^S in PhoB mutants	Mutation in rssB	Restored accumulation of σ^S in phoB, rssB mutants	RssB and PhoB manage σ^S accumulation	*E. coli*	Tu, 2006
To understand why σ^{32} mutants cannot overcome heat shock	σ^A-285	Heat shock can be overcome in σ^{32}, σ^A-285 mutants	σ^{32} has a lower binding affinity for core RNAP	*E. coli*	Zhou, 1992
To understand why AlgU (σ^H) deregulation leads to mucoidy	Mutants in mucA and mucB	MucA and MucB mutants become mucoid	MucA and MucB regulate AlgU control of mucoidy	*P. aeruginosa*	Martin, 1993

[a]Reason or investigation conducted in the study or mutant being investigated

[b]Suppressor mutation identified that restored mutant phenotype

[c]Phenotype of the suppressor mutant

[d]Implication or conclusion made from analyzing suppressors

[e]Organism in which study was performed

(Tu et al. 2006). The implications from these results is that suppressor mutations are able to reveal previously unknown regulators of alternative σ factors. Performing a similar study searching for suppressors of *comW* mutants in *S. pneumoniae* will answer questions that remain about the function of ComW in σ^X activity and its role during transformation.

1.6 The small protein ComW is necessary for σ^X activity, stability, and the full competence response.

Despite the absolute reliance on σ^X for competence, the full transformation response is not achieved without the cooperation of another protein, ComW. ComW was first identified as a CSP-inducible open reading frame, which upon deletion, was shown to be required for transformation (Bartilson et al. 2001). At this time, ComW sequences were only identified in *S. pneumoniae* and its closely relates species, *S. mitis*. However, with the availability of more streptococcal genomes, ComW sequences have now identified in a total of eleven related species of streptococci (*S. pneumoniae*, *S. mitis*, *S. pseudopneumoniae*, *S. oralis*, *S. cristatus*, *S. oligofermentans*, *S. anginosus*, *S. infantis*, *S. dentisani*, *S. tigurinus*, and *S. sinensis*). The only proteins that show significant similarity to *S. pneumoniae* ComW are ComW sequences in other streptococcal species. Amongst the species, similarity to the *S. pneumoniae* ComW ranges from 100% in other pneumococci to 40% in *S. tigurinus* (Fig. 7). There are no domains in ComW that have homology to any other types of identified domains of known function. However, the alignment of ComW sequences from all available species reveals three conserved regions that may yield future clues to ComW function.

ComW is an early gene with similar temporal induction patterns to σ^X (Bartilson et al. 2001; Peterson et al. 2004) and is implicated in being responsible for the full σ^X response during

```
                                                                                  % Identity
Species                                          Sequence                          to S.pneumo

S.pneumo      MIMLQKIYEQMANFYDSIEEEYGPTFGDNFDWEHVHFKFLIYYLVRYGIGCRKDFIVYHYRVAYRLYLEKLVMNRGFISC    -
S.pseudo      --MLQSIYEQMTDFYRNIEEEYGTVFGDNFDWEHFHFKFLIYYLVRYGIGCRRDFIVYHYRVAYRLYLEKMIMNRGFISC    86
S.mitis       --MLQNIYDQMTDFYDSIEEEYATFFGNSWDWEHFHFKFLIYYLVRYGIGCRRDFIVYHYRVAYRLYLEKMIMKQGFISC    79
S.oralis      --MIEQIYEQYLDFYDVIEKEYSYLVDNDLEWEVFHLRFLLYYLVRYKLDIMHPLFSYHYRACYRLYIEQLLISNDWVGG    43
S.crista      MMEIQEIYNQFRDYYGELEAEYAHCQKASTEWESLHLRYLIYYLMRYGIGEMKFFNAYHYRAAYRWYLQSLMLSSA----    45
S.oligof      -MEIQEIYNQFRDYYGELEAEYAHCQKASMEWESLHLRYLIYYLMRYGIGEMKFFNAYHYRAAYRWYLQSLMLSST----    46
S.angino      -MEIQEIYNQFRDYYGELEAEYAHCQKASMEWESLRLHYLIYYLIRYGIGEMKFFNAYHYRAAYRWYLQSLMLSSA----    49
S.sinens      -MKIQEIYLKYKGYYAEIEAEYSHCKKTSIEWETLHLRYLIYYLVRYNIGKMQFFNPYHYRTAYRLYLEQLVAS------    44
S.infant      --MIEQIYEQYLDFYDVIEKEYSYLVDNDLEWEVFHLRFLLYYLVRYKLDTMHPLFSYHYRACYRLYIEQLLISNDWVGR    43
S.dentis      --MLQKFYERAFVFLKLVEQEYVSLCQSSAEWESLHLRFLLYYLIRFKIKSDRDFSLYHFKTAYRLYLDKFLQGGTTLIQ    44
S.tiguri      --MLQKFYDRAFVFLKLVEQEYASLGQSCAEWESLHLRFLLYYLIRFKIKSDRDFSLYHFKTAYRLYLDEFLQGGTTLIQ    40
                 ::.:* :   :   :* **        :** .::::*:***:*: :   : :  **::..** *::.::
```

Figure 7: Alignment of ComW homologs from *S. pneumoniae* and all other species of streptococci with identifiable ComW sequences. *S. pneumoniae* ComW residues (accession number AAK98824.1) were aligned with corresponding residues of ComW sequences from *S. pseudopneumoniae* (WP_049538978.1), *S. mitis* (WP_000940045.1), *S. oralis* (WP_049503840.1), *S. cristatus* (WP_005591530.1), *S. oligofermentans* (WP_015605924.1), *S. sinensis* (WP_037617413.1), *S. anginosus* (EJP26452.1), *S. infantis* (EF053736.1), *S. dentisani* (WP_038804352.1), and *S. tigurinus* (EMG33941.1). Alignment was made using Clustal Omega with default parameters (Sievers et al. 2011). Species name is on left, amino acid sequence is in middle, percent sequence identity to *S. pneumoniae* sequence is on right. Asterisks, identical residues; colons, conserved residues; periods, semi-conserved residues.

competence (Luo et al. 2004). In the latter study, Luo showed that ectopic expression of both ComW and σ^X was sufficient to restore transformation to 80% of WT (wild type) levels. Next, ComW was shown to be required for the full competence response, σ^X activity, and σ^X stability (Sung & Morrison 2005). Several observations about ComW have been made in *comW* mutants. First, the amount of transformants produced from a *comW* mutant culture is severely reduced, 10^4- to 10^5-fold below the level of WT (Luo et al. 2004; Bartilson et al. 2001). Second, *comW* mutants display a reduced amount of late gene transcription, about 10% of the level measured in WT (Sung & Morrison 2005). Third, the amount of σ^X protein is also 10% of the WT level when measured at 20 minutes after CSP induction, the time of peak σ^X activity (Sung & Morrison 2005). Even when the amount of σ^X is stabilized in *clpE* mutants, where σ^X is not degraded by the ClpEP protease, the amount of transformants remains at the *comW* mutant level, suggesting ComW is also required for σ^X activity (Sung & Morrison 2005). Finally, the N- and C- termini are required for σ^X stability and activity. More specifically, both termini are required for σ^X stability while the N-terminus is necessary for both σ^X stability and activity (Sung & Morrison 2005). Despite the data available about the many functions of ComW during competence, the mechanism of ComW's activity remains unclear.

1.7 Specific Aims

Over the eighty-eight years since the first observation of transformation in *S. pneumoniae*, much of the competence mechanism has been elucidated. However, despite extensive study into the induction of competence and the discovery of the central role of σ^X, some aspects still remain unclear. The small protein, ComW, is critical for the full transformation response in *S. pneumoniae*, and studies attempting to understand its function have produced several observations about its regulation of σ^X. However, despite the proposed roles for ComW, its exact responsibility

in σ^X regulation during competence is not fully understood. Because direct attempts to study ComW had not provided a clear model of function, we decided to explore ComW function using an unbiased, indirect approach.

This study first aimed to use a suppressor screen to identify suppressor mutations that bypass the need for ComW during transformation. The identity of such a suppressor could provide a clue as to the role of ComW during the transformation process. Spontaneous mutations as the source of suppressor mutations rather than chemical mutagenesis or transposon insertions requires whole genome sequencing to locate the suppressor. Once the suppressor was identified, this study aimed to confirm the linkage of suppression to the mutation found during whole genome sequencing. Then, this study aimed to understand the function of the mutated protein, and how it may play a role with ComW during transformation.

Next, this study aimed to investigate the mechanism of transformation restoration due to the suppressor mutation. Transformation efficiency is a complicated response that requires the coordination of all cells to respond and correctly express the necessary transformation specific genes and proteins. While transformation restoration by the suppressor suggests a role for ComW that also implicates that suppressor, the exact mechanisms of restoration of transformation need to be investigated, such as restoration of late gene expression or restoration of the amount of σ^X protein, or possibly both. This sheds light on the function of the suppressor in bypassing the ComW requirement and the overall role for ComW during transformation.

Lastly, this study aimed to explore the general transformation efficiency of a wild type strain when transformed with genomic and PCR amplified DNA. Genetic manipulation in *S.* pneumoniae usually involves the use of an inserted selectable marker, potentially having polar effects on gene expression. We wanted to use the high natural transformability of *S. pneumoniae*

to explore the range of transformation efficiency when subjected to a range of DNA donors. This allows for establishing a set of parameters to use during genetic manipulation to maximize strain creation for genetic studies. Then, these parameters were used to perform markerless genetic manipulations to minimize disruption in the genome.

This study aimed to explore the transformation process of *S. pneumoniae*, but more specifically, it aimed to explore the role of the small protein ComW. Understanding the function of ComW in the regulation of competence could reveal another layer of control of the alternative σ factor, or it may reveal a novel mechanism of σ^X regulation.

II. MATERIALS AND METHODS

(Portions of this chapter were published in: Tovpeko, Y. and Morrison, D.A. Competence for Genetic Transformation in *Streptococcus pneumoniae*: Mutations in σ^A Bypass the *comW* requirement. Journal of Bacteriology. 196: 3724-3734, 2014)

2.1 Bacterial strains and culture media

The strains used in this study are listed in Table II. The primers used in this study are listed in Table III. CP2137, a Δ*cps* Δ*comA* derivative of strain Rx1, was used as the WT standard for transformation assays. All strains were cultured in CAT medium, supplemented as needed with 1.5% agar. CAT medium was prepared from 5 g of tryptone (Difco Laboratories), 10 g of enzymatic casein hydrolysate (ICN Nutritional Biochemicals), 1 g of yeast extract (Difco), and 5 g of NaCl in 1 liter of H_2O; sterilized for 40 min at 121°C; and then supplemented to 0.2% glucose and 0.016 M K_2HPO_4 (CAT+GP) before use. Antibiotics were used at the concentrations listed in Table IV. CSP1 (Pozzi et al. 1996) was obtained from NeoBioSci (Cambridge, MA) as a custom synthetic peptide with sequence EMRLSKFFRDFILQRKK at 80% purity and stored as sterile 0.025% solution in water at −20°C.

2.2 Sanger sequencing

For verification of the sequences of new strains or PCR amplicons, regions of interest were amplified as 5- to 7-kb amplicons with the following conditions. Amplification was performed in 50-µl reaction mixtures with 1 µl Phire HotStart II polymerase (Thermo Scientific), 1X Phire Reaction Buffer, 10 ng template DNA, 200 µmol/liter of each deoxynucleoside triphosphate (dNTP), and 0.5 µmol/liter of each primer. The amplification conditions were 98°C for 90 s and then 98°C for 15 s, 56°C for 15 s, and 72°C for 105 s for 30 cycles, followed by a 5-min 72°C final extension. PCR products were purified using a DNA Clean and Concentrate-5 kit (Zymo

TABLE II. Bacterial strains used in this study.

Strain	Description	Source*
S. pneumoniae		
CPM7	CP1250, but ssbB−::{pEVP3}::ssbB+; SsbB+; Sm^R Cm^R	(Lee & Morrison 1999)
CP1250	Rx1, but *hex malM511 str-1 bgl-1;* low α-galactosidase background; Sm^R	(Pestova et al. 1996)
CP1344	CP1250, but Δ*clpC*::PcTet; Sm^R Tc^R	(Luo et al. 2004)
CP1376	CP1250, but *comW*::*kan*; Sm^R Kn^R	(Sung & Morrison 2005)
CP1500	Rx1, but *hex nov-r1 bry-r str-1 ery-r1 ery-r2;* Nv^R Em^R Sm^R	(Cato & Guild 1968)
CP1759	CP1250, but *comW*::Spc; Spc^R Sm^R	(Luo et al. 2004)
CP2000	CP1250, but Δ*cps;* Sm^R	(Weng et al. 2013)
CP2107	CP2000, but Δ*cps,* ssbB−::{pEVP3}::ssbB+; SsbB+; Sm^R Cm^R	CP2000 X CPM7 (Weng et al. 2013)
CP2136	CP2107, but *comA::cheshire*; Sm^R Cm^R Em^R	(Weng et al. 2009)
CP2137	CP2000, but Δ*cheshire, comA*$^-$; Sm^R Cm^R	(Weng et al. 2009)
CP2451	CP2137, but *rpoD*-L363F; Sm^R Cm^R	CP2137 X NYT1 {rpoD}
CP2452	CP2137, but *rpoD*-A171V; Sm^R Cm^R	CP2137 X ALT4 {rpoD}
CP2453	CP2137, but *rpoD*-R355H; Sm^R Cm^R	CP2137 X FLT4 {rpoD}
CP2454	CP2137, but *rpoD*-R316H; Sm^R Cm^R	CP2137 X ILT1 {rpoD}

* { }, transfer of gene only via PCR amplicon

X, transformation cross as recipient X donor

[1] Followed by transformation with *nov-r1, ery-r1 ery-r2*, and Δ*clpC*::PcTet donors

TABLE II. (continued)

Strain	Description	Source*
CP2455	CP2451, but Δ*comW::kan*; Sm^R Cm^R Kn^R	CP2451 X CP1376 {ΔcomW::kan}
CP2456	CP2452, but Δ*comW::kan*; Sm^R Cm^R Kn^R	CP2452 X CP1376 {ΔcomW::kan}
CP2457	CP2453, but Δ*comW::kan*; Sm^R Cm^R Kn^R	CP2453 X CP1376 {ΔcomW::kan}
CP2458	CP2454, but Δ*comW::kan*; Sm^R Cm^R Kn^R	CP2454 X CP1376 {ΔcomW::kan}
CP2459	ALT4, but *rpoD*-V171A; Sm^R Cm^R Kn^R Nv^R Em^R Tc^R Kn^R	ALT4 X CP2137 {rpoD}
CP2460	FLT4, but *rpoD*-H355R; Sm^R Cm^R Kn^R Nv^R Em^R Tc^R Kn^R	FLT4 X CP2137 {rpoD}
CP2461	ILT1, but *rpoD*-H316R; Sm^R Cm^R Kn^R Nv^R Em^R Tc^R Kn^R	ILT4 X CP2137 {rpoD}
CP2462	NYT1, but *rpo*D-F363L; Sm^R Cm^R Kn^R Nv^R Em^R Tc^R Kn^R	NYT1 X CP2137 {rpoD}
CP2463	CP2137, Δ*comW::kan*; Sm^R Cm^R Kn^R	CP2137 X CP1376 {ΔcomW::kan}
CP2464	CP2137, Δ*comW::kan,* but *rpoD*-region 2-4 chimera from *S. mutans*; Sm^R Cm^R Kn^R	CP2137 X {*rpoD*-chimera} X CP1376 {Δ*comW::kan*}
ALT1	CP2463, but *rpoD*-P287S; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
ALT2	CP2463, but *rpoD*-R316C; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
ALT3	CP2463, but *rpoD*-A171V; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
ALT4	CP2463, but *rpoD*-A171V; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
ALT5	CP2463, but *rpoD*-R297C; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]

* { }, transfer of gene only via PCR amplicon

X, transformation cross as recipient X donor

[1] Followed by transformation with *nov-r1, ery-r1 ery-r2*, and Δ*clpC*::PcTet donors

TABLE II. (continued)

Strain	Description	Source*
DET3	CP2463, but *rpoD*-A171V; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
DET5	CP2463, but *rpoD*-A171V; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
FLT1	CP2463, but *rpoD*-R316H; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
FLT3	CP2463, but *rpoD*-R355H; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
FLT4	CP2463, but *rpoD*-R355H; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
FLT5	CP2463, but *rpoD*-R355H; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
GAT1	CP2463, but *rpoD*-R316H; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
GAT2	CP2463, but *rpoD*-R316H; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
GAT3	CP2463, but *rpoD*-A171V; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
GAT4	CP2463, but *rpoD*-R316H; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
ILT1	CP2463, but *rpoD*-R316H; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
ILT2	CP2463, but *rpoD*-R316H; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
ILT4	CP2463, but *rpoD*-R316H; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
ILT5	CP2463, but *rpoD*-R316H; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]

* { }, transfer of gene only via PCR amplicon

X, transformation cross as recipient X donor

[1] Followed by transformation with *nov-r1, ery-r1 ery-r2*, and Δ*clpC*::PcTet donors

TABLE II. (continued)

Strain	Description	Source*
MAT3	CP2463, but *rpoD*-R316H; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
MAT4	CP2463, but *rpoD*-Y290C; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
MAT5	CP2463, but *rpoD*-A171V; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
MTT1	CP2463, but *rpoD*-R316H; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
MTT2	CP2463, but *rpoD*-A171V; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
MTT3	CP2463, but *rpoD*-A171V; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
MTT4	CP2463, but *rpoD*-R316H; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
MTT5	CP2463, but *rpoD*-R316H; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
NCT2	CP2463, but *rpoD*-A171V; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
NCT5	CP2463, but *rpoD*-A171V; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
NVT1	CP2463, but *rpoD*-R316H; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
NVT2	CP2463, but *rpoD*-V314A; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
NVT3	CP2463, but *rpoD*-V314A; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
NVT4	CP2463, but *rpoD*-A171V; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]

* { }, transfer of gene only via PCR amplicon

X, transformation cross as recipient X donor

[1] Followed by transformation with *nov-r1, ery-r1 ery-r2*, and Δ*clpC*::PcTet donors

TABLE II. (continued)

Strain	Description	Source*
NVT5	CP2463, but *rpoD*-A171V; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
NYT1	CP2463, but *rpoD*-L363F; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
NYT2	CP2463, but *rpoD*-R316H; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
NYT3	CP2463, but *rpoD*-R316H; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
NYT4	CP2463, but *rpoD*-L363F; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
OHT1	CP2463, but *rpoD*-R355H; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
PRT3	CP2463, but *rpoD*-R355H; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
PRT4	CP2463, but *rpoD*-R316H; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
PRT5	CP2463, but *rpoD*-R316H; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
PRT6	CP2463, but *rpoD*-R316H; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
SCT2	CP2463, but *rpoD*-R316H; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
SCT4	CP2463, but *rpoD*-R316H; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
SCT5	CP2463, but *rpoD*-R316H; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
TXT2	CP2463, but *rpoD*-R316H; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]

* { }, transfer of gene only via PCR amplicon

X, transformation cross as recipient X donor

[1] Followed by transformation with *nov-r1, ery-r1 ery-r2*, and Δ*clpC*::PcTet donors

TABLE II. (continued)

Strain	Description	Source*
TXT4	CP2463, but *rpoD*-R316H; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
TXT5	CP2463, but *rpoD*-R316H; Sm^R Cm^R Nv^R Em^R Tc^R Kn^R	Spontaneous *rpoD* mutation[1]
S. mutans		
UA159	Transformable S. mutans reference strain	(Tao et al. 1993)

* { }, transfer of gene only via PCR amplicon

X, transformation cross as recipient X donor

[1] Followed by transformation with *nov-r1, ery-r1 ery-r2*, and Δ*clpC*::PcTet donors

TABLE III. Primers used in this study.

Primer	Primer {5'-3'}	Location
S. pneumoniae		
DAM497	CAATTGACTATATTAGAGGCGAGACA	*transposase_B*
DAM500	TATCAAGCGCATCATTCAAGATAACAG	*purA*
YT18	GCCCTAGAAGAATTGGAACG	*dnaG*
YT20	CAGGGTCGATGACTTCTTCC	*spr0980*
YT30	GACAGGCTTTGAGTCTCTTGATGG	*ABC-NBD*
YT31	CGGACGCTCAAACTTGGCTAATTC	*spr0982*
YT34	CCATTGCCAAACGCTATGTC	*rpoD*
YT36	GTCAACCGCCTTCATCAAGCC	*rpoD*
YT40	TATGGGCTTGATGAAGGC	*rpoD*
YT41	TATGGGCTTGATGAAGGT	*rpoD*
YT42	ATACGCTCACGAGTTACG	*rpoD*
YT43	CGTGAAGAAAATGTTCTGCG	*rpoD*
YT44	CGTGAAGAAAATGTTCTGCA	*rpoD*
YT45	TCACATCTGCCTCGATTG	*cpoA*
YT46	TTGCTACGACTTGGTTGGC	*rpoD*
YT47	TTGCTACGACTTGGTTGGT	*rpoD*
YT48	GTCAATGACCCTGTCCGTATG	*cpoA*
YT49	CAAGTCGTAGCAAACCGC	*rpoD*
YT50	CAAGTCGTAGCAAACCGT	*rpoD*
YT51	CACGGTAAGCACCTGAAAC	*cpoA*
YT66	AAAGGAGCCAAGTTAAGCAAATGACAGC	*spr0976*
YT67	GGAGAGCAGTGCCATATTTCCTTCTTGAATC	*rpoD*
YT68	GCGCAGCAGTGGAAAATATGGCTTATACAG	*spr0980*
YT69	CCGTAACAGCAAGTAAGAACACCCCAAAT	*spr0983*
YT76	AGCGCCGACAGGGATTGGGA	*dinG*
YT77	ACATTGGCCTTTTGACGTGCAT	*ezrA*
YT102	GGGTTCCTGCAATTCAGTTAG	*rplC*
YT103	GCAACTGGGCCAAAGATAC	*rplW*
YT106	CACTTGAAGATGTGGGGAAAGTCTTTA	*rpoD*
YT107	GCACCTACTACGATAAACTGATTGTCAC	*rpoD*
YT108	CGCTTGAAGATGTTGGTAAAGTCTTTG	*rpoD*
S. mutans		
YT104	GCCGAGCAGTGGGGTTTGATGAAGGCAGTTG	*rpoD*
YT105	CGCCGGCAGTGTCCTCTCTTAATCTTCAACAAAATC	*SMU_823*

TABLE IV. Antibiotic concentrations used in this study.

		Concentration (μg/mL)	
Antibiotic	Abbreviation	Overlay[a]	Final/Liquid[b]
Erythromycin	Em	0.3	0.075
Kanamycin	Kn	800	200
Novobiocin	Nv	10	2.5
Spectinomycin	Spc	400	100
Tetracycline	Tc	1	0.25
Trimethoprim	Tr	400	100

[a]Fourth layer agar overlay in sandwich plating (Sung & Morrison 2005)

[b]Concentration used in liquid media or in single layer plates

Research), and recovered in TE buffer. The concentration of purified products was measured using a NanoDrop 1000 Spectrophotometer (Thermo Scientific). Products were stored at -20°C. 10-μL DNA samples to be sequenced were prepared at 10 ng/μL in H_2O and primers for sequencing were prepared at 5 mM in H_2O. At least one forward and reverse primer was used to ensure at least 2-fold coverage of the region of interest. Samples and primers were sequenced by the University of Illinois at Chicago (UIC) Research Resources Center (RRC) Sequencing Center. Sequences were analyzed using Clone Manager (Sci-Ed Software).

2.3 Strain construction and verifications

2.3.1 Construction of strain CP2463

To construct a Δ*comW* mutant strain, a 5.5-kb region surrounding the Δ*comW::kan* locus from CP1376 (Sung & Morrison 2005) was amplified using primers DAM497 and DAM500. PCR products were purified using a DNA Clean and Concentrate-5 kit (Zymo Research), recovered in TE buffer, and transformed into CP2137. After the standard transformation and plating on kanamycin (section 2.5), 10 individual colonies were isolated, outgrown to OD_{550} 0.2, and frozen with 10% glycerol. Thawed cells were diluted 1:100 into CAT+GP and kanamycin, and grown to OD_{550} 0.2. Individual clones were tested for correct integration of Δ*comW::kan*. Positive clones were streaked on kanamycin plates, and a subclone was tested for correct integration of the fragment, and then verified by Sanger sequencing of the Δ*comW::kan* locus and surrounding 500 base pairs (bp). This positive subclone was retained and named CP2463 (Fig. 8).

2.3.2 Isolation of spontaneous mutant strains ALT1 – TXT5

To identify spontaneous mutants that restore transformation efficiency in a Δ*comW* mutant strain, 14 individual colonies of CP2463 were outgrown to OD_{550} 0.05 and transformed, then recovered as enrichment libraries as described in section 2.7. Each of the 14 parallel outgrowths

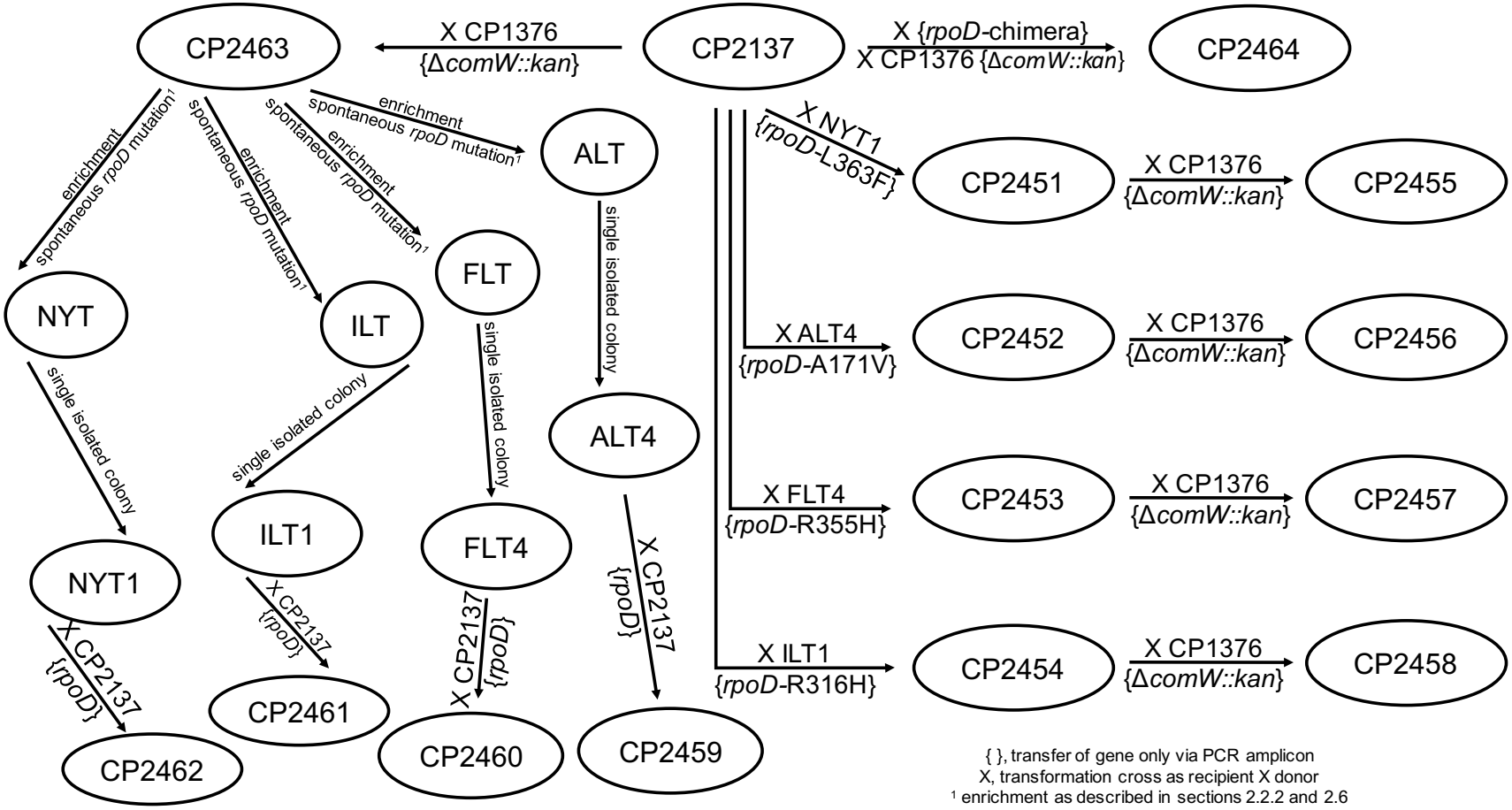


Figure 8. Pedigree of strains created in this study. Schematic of strains created and transformations performed to create the strains in this study. Each strain is represented by a circle, and arrows coming from the strain indicate the parent strain, arrows pointing to the strain indicate the new strain. Text across the arrow explain the transformation performed in order to create that strain. Strain genotypes and descriptions can be found in Table I. Circles, strains; arrows, action performed to create strain; { }, transfer of only bracketed gene from the donor strain by PCR amplicon; X, transformation as recipient X donor; [1], spontaneous *rpoD* mutation identified after serial enrichment as described in sections 2.3.2 and 2.7.

of the Δ*comW* mutant was given a two-letter code: AL, DE, FL, GA, IL, MA, MT, NC, NV, NY, OH, PR, SC, or TX. The name of each parallel enrichment was appended with each subsequent transformation. After the first enrichment with Nv, collected Nv^R pools were named with -N at the end of their two-letter code. For example, the AL spontaneous mutant pool transformed with Nv^R donor DNA and the collected Nv^R pool was named ALN. After the second enrichment with Em, collected Em^R pools were named with –E at the end of their two-letter code. For example, ALN pools were transformed with Em^R donor DNA and collected Em^R pools were named ALE. After enrichment with Tc, collected Tc^R pools were named with –T at the end of their two-letter code. For example, ALE pools were transformed with Tc^R donor DNA and collected Tc^R pools were named ALT. At the end of the enrichments each Tc^R pool was plated on CAT agar with Tc, and 5 colonies from each of the 14 parallel enrichments were isolated, outgrown to OD_{550} 0.2, and frozen with 10% glycerol. Each colony was given a number from 1 through 5, for example, the third colony isolated at the end of the NYT parallel enrichment was named NYT3 (Fig. 8). Whole genome sequences were obtained (as described in section 2.8) for one isolated strain from each parallel enrichment: ALT4, DET5, FLT4, GAT1, ILT5, MAT5, MTT5, NCT2, NVT1, NYT1, NYT2, OHT1, SCT4, and TXT4 (see Table S1 in Appendix). The *rpoD* mutations in each of the 5 clones from each of the 14 enrichments were identified by amplifying the *rpoD* locus of each isolated clone using primers YT18 and YT20, purifying the PCR products using a DNA Clean and Concentrate-5 Kit (Zymo Research), and using Sanger sequencing.

2.3.3 Construction of strains CP2451 – CP2454

To construct specific backcross *rpoD* mutant strains in the CP2137 background, the *rpoD* loci from strains NYT1 (L363F), ALT4 (A171V), FLT4 (R355H), and ILT1 (R316H) were amplified using primers YT30 and YT31, as described in section 2.11. The 5.5-kb PCR products

were purified using a DNA Clean and Concentrate-5 Kit (Zymo Research) and transformed at 100 ng/mL into CP2137 separately for each different *rpoD* amplicon. After the standard transformation and plating (section 2.4), 10 individual colonies from each transformation were isolated, outgrown to OD_{550} 0.2, and frozen with 10% glycerol. Thawed cells were diluted 1:100 into CAT+GP, and grown to OD_{550} 0.2. Cells were plated on CAT agar. After overnight incubation at 37°C, individual colonies were stabbed into CAT+GP, outgrown to OD_{550} 0.2, and then frozen with 10% glycerol. The *rpoD* sequence was verified by the mismatched-primer assay as described in section 2.11 and 2.12. 1 μL of thawed cells was used as template for PCR amplification of the *rpoD* region as described above. PCR products were purified and submitted for Sanger sequencing as described above. Positive colonies were re-streaked, and single subclones were again tested for correct integration. Single purified subclones of one positive colony for each strain was verified by Sanger sequencing of the entire *rpoD* locus. The positive subclones were retained and named CP2451 (σ^A-L363F), CP2452 (σ^A-A171V), CP2453 (σ^A-R355H), and CP2454 (σ^A-R316H) (Fig. 8).

2.3.4 Construction of strains CP2455 – CP2458

Due to the low transformation efficiency of the Δ*comW* mutant, markers cannot be introduced into the Δ*comW* mutant background. Instead, the marker must first be introduced into the WT background, and then *comW* can be deleted from the strain. To construct *rpoD* mutant strains in the CP2463 (CP2137, but Δ*comW::kan*) background, strains CP2451, CP2452, CP2453, and CP2454 were transformed separately with the Δ*comW::kan* donor region from CP1376 (Sung & Morrison 2005). After the standard transformation and plating on kanamycin (section 2.4), 10 individual colonies from each transformation were isolated, outgrown to OD_{550} 0.2, and frozen with 10% glycerol. Thawed cells were diluted 1:100 into CAT+GP, grown to OD_{550} 0.2, and then plated on CAT. After overnight incubation at 37°C, individual colonies were stabbed into

CAT+GP, outgrown to OD_{550} 0.2, and then frozen with 10% glycerol. 1 μL of thawed cells was used as template for PCR amplification of the *rpoD* region as described above. PCR products were purified sequenced as described above. Positive clones were re-streaked, and single subclones were again tested for correct integration. A single subclone of one positive colony for each strain was verified by Sanger sequencing of the Δ*comW::kan* locus and the surrounding 500 bp. The positive subclones were retained and named CP2455 (σ^A-L363F), CP2456 (σ^A-A171V), CP2457 (σ^A-R355H), and CP2458 (σ^A-R316H) (Fig. 8).

2.3.5 Construction of strains CP2459 – CP2462

To construct spontaneous mutant strains that retained all mutations acquired during enrichment except *rpoD*, the WT *rpoD* locus from CP2137 was amplified using primers YT30 and YT31. PCR products purified as described above, and transformed separately into ALT4, FLT4, ILT1, and NYT1. After the standard transformation and plating (section 2.4), 10 individual colonies from each transformation were isolated, outgrown to OD_{550} 0.2, and frozen with 10% glycerol. Thawed cells were diluted 1:100 into CAT+GP, grown to OD_{550} 0.2, and plated on CAT agar. After overnight incubation at 37°C, individual colonies were stabbed into CAT+GP with kanamycin, and outgrown to OD_{550} 0.2, then frozen with 10% glycerol. The *rpoD* sequence was verified by the mismatched-primer assay as described in sections 2.11 and 2.12. 1 μL of thawed cells was used as template for PCR amplification of the *rpoD* region as described above. PCR products were purified and submitted for Sanger sequencing as described above. Positive defined clones were re-streaked, and single subclones were again tested for correct integration. A single purified subclone of one positive colony for each strain was verified by Sanger sequencing of the entire *rpoD* locus. The positive subclones were retained and named CP2459 (σ^A-V171A), CP2460 (σ^A-H355R), CP2461 (σ^A-H316R), and CP2462 (σ^A-F363L) (Fig. 8).

2.4 Preparation of donor DNA

The primers used for PCR amplification of donor DNA are listed in Table II. The *comW*::*kan* deletion fragment was amplified from strain CP1376 using primers DAM497 and DAM500. *rpoD* fragments (5.5 kb) were amplified from strain CP2137 using primers YT30 and YT31. Amplification was performed in 50-µl reaction mixtures with 1 µl Phire HotStart II polymerase (Thermo Scientific) and Phire Reaction Buffer, 10 ng template DNA, 200 µmol/liter of each deoxynucleoside triphosphate (dNTP), and 0.5 µmol/liter of each primer. The amplification conditions were 98°C for 90 s and then 98°C for 15 s, 56°C for 15 s, and 72°C for 105 s for 30 cycles, followed by a 5-min 72°C final extension. The *gyrB* Nv^r marker was prepared as a 7.4-kb amplicon using primers YT76 and YT77 and CP1500 DNA as the template. Amplification was performed as described above, except with 65°C as the annealing temperature. The 50-µL PCR products were purified using a DNA Clean and Concentrate kit (Zymo Research). Genomic donor DNA was extracted from strains CP1500, CP1344, and CP1759 as previously described (Sung & Morrison 2005) (Table II).

2.5 Transformation assays

The standard assay for transformation was done essentially as previously described (Sung & Morrison 2005). A log-phase culture at an optical density at 550 nm (OD_{550}) of 0.05 at 37°C was incubated with 0.1 µg/ml DNA, 250 ng/ml CSP, 0.5 mM $CaCl_2$, and 0.04% bovine serum albumin for 80 min at 37°C. Portions of the culture were then embedded (in 1.5 ml of CAT mixed with 1.5 ml of CAT agar) in sandwich plates and overlaid with the relevant antibiotic, as previously described (Sung & Morrison 2005). After 15 h at 37°C, colonies were counted. The transformation efficiency was expressed as CFU per ml of cells transformed at an OD_{550} of 0.05. To determine the transformation efficiency relative to the WT, the transformant yield of a strain was divided by

that for a parallel WT culture. In the WT, typical yields for the PCR-amplified Nv^r donor were 50% to 90% of cells transformed, while typical yields for the genomic DNA preparations were ~2%, ~1%, and 0.1 to 0.3% for Em^r, Tc^r, and Spc^r markers, respectively. *gyrB* (Nv^R) and *rplD* (Em^R) markers were prepared as 7.4-kb amplicons. PCR products were purified using a DNA Clean and Concentrate kit (Zymo Research).

2.6 Estimation of spontaneous mutation frequency

An indication of the number of spontaneous mutations present in a mutant library was obtained by determining the number of trimethoprim-resistant mutants, which arise by inactivation of any of several proteins, including thymidine synthetase and the Ami transporters. A single CP2137 colony was isolated and outgrown to OD_{550} 0.2 and frozen with 15% glycerol. Thawed cells were diluted in CAT+GP and grown for 6 generations (~4 hours). 100 μL aliquots were plated on sandwich plates with trimethoprim as previously described (Sung & Morrison 2005). Mutants insensitive to 100 to 200 μg/ml trimethoprim occurred at a frequency of ~10^{-4}.

2.7 Transformation and recovery of enrichment libraries

Mutant libraries were prepared by serial dilutions of cells from a single isolated colony of CP2463 in culture volumes of 10 ml. Every 4 hours, approximately 6 doublings, the culture was diluted 1:100 into fresh CAT+GP for a total of 35 doublings over a period of 24 hours. After 24 hours, the cells were frozen at OD_{550} 0.2 in 10% glycerol. To obtain an enriched suppressor mutant library, 10^8 CFU of the above thawed cells in 1 ml CAT were exposed to CSP and DNA as described for the standard transformation assay. The entire culture was spread onto 20 100-mm selective CAT agar plates. After 15 h at 37°C in 5% CO_2, there were ~1000 colonies per plate. The 2×10^4 resulting transformant colonies were combined by resuspension in 10 ml CAT. The resuspension was diluted 1:10 into fresh CAT+GP. During outgrowth at 37°C for 6-7 hours, viable

cells increased from ~10^5 to 10^8 CFU/ml. The starting OD_{550} of the diluted resuspension was ~0.02 and the cells were harvested at 0.2. This pool of transformant clones was stored at −80°C with 10% glycerol. The recovered transformant library pool is estimated to contain 20,000 transformant clones, each at 5,000 cells/ml.

2.8 Whole-genome sequencing (WGS)

Genomic DNA for sequencing was prepared from 10^9 CFU by 0.2% Triton lysis and purified using a Genomic DNA Clean and Concentrate-25 kit with 25 μg capacity (Zymo Research). For sequencing on the Ion Torrent Proton sequencer, genomic DNA extracts were sheared at the RRC to an average fragment size of 200 bp using a Covaris S220 focused ultrasonicator and purified using AMPure beads (Beckman-Coulter, Brea, CA). Library preparation followed the Wafergen Prep PGM 200 DNA Library protocol, using an Apollo 324 system (Wafergen, Fremont, CA) according to the manufacturer's instructions. DNA libraries containing Ion Xpress Barcode Adapters (Life Technologies, Grand Island, NY) were quantified in a Qubit 2.0 fluorometer (Life Technologies) and pooled in approximately equimolar ratio. The pool was quantified using the Library Quantification kit for Ion Torrent (KAPA Biosystems, Woburn, MA) and diluted to 26 nM prior to emulsion PCR, which was performed on the OneTouch2 instrument (Life Technologies) according to the manufacturer's instructions. Subsequently, the emulsion was broken, and Ion sphere particles (ISPs) containing amplified DNA fragments were recovered using the OneTouchES instrument (Life Technologies). The ISPs were loaded onto a Proton I chip and analyzed on the Ion Torrent Proton sequencer using 200-bp chemistry (V2) according to the manufacturer's instructions. Sequence data were demultiplexed on the Proton server and exported as FASTQ files. Using the CLC genomics workbench (CLC Bio), sequencing reads from each strain were mapped to the reference genome sequence of strain

R6 (Hoskins et al. 2001), a laboratory ancestor of Rx1 (Tiraby & Fox 1973). We obtained an average of 150X coverage for each strain, at least 95% of each read matching the reference. Variants that indicated a single nucleotide polymorphisms (SNP) in at least 98% of the reads that mapped to that region in any of the strains relative to the reference were compiled as an Excel Spreadsheet (Microsoft Office) and filtered as described in section 2.9.

2.9 Whole-genome sequencing analysis

To identify SNPs resulting in non-synonymous amino acid changes that conferred the suppressor phenotype, we analyzed all SNPs collected in WGS. In 14 sequenced strains, 9000 total SNPs were identified, approximately 600 per strain. SNPs in regions with less than 40X coverage and not mapping to open reading frames were removed from consideration. Of the remaining ~6,600 SNPs, ~2,700 resulted in synonymous amino acid changes and were removed from consideration. 3,768 of the remaining ~3,900 SNPs were variations caused by divergence of the Rx1 and R6 lineages (Tiraby & Fox 1973), and because they were present in every single strain, they were also removed from consideration. 64 SNPs mapped to donor drug markers introduced during enrichment, and were also removed from consideration. 68 SNPs remained that were distributed among 53 unique genes, 2 to 10 SNPs per strain (Table S1 in Appendix).

2.10 Targeted gene sequencing at the *rpoD* locus

To sequence DNA near the *rpoD* gene, a 5.5-kb amplicon was prepared using primers YT30 and YT31 as described above, followed by sequencing reads using primers YT18, YT20, YT34, and YT36 at the University of Illinois at Chicago (UIC) Research Resources Center (RRC) sequencing facility, providing >3-fold coverage of 100% of the *rpoD* gene. The reads were generally ~1000 bases, and identical to the expected sequences except at the bases being investigated. Sequence data was unreliable for the first 50 bases and the last 100 bases of the

sequence, so any identified SNPs in these regions these were not regarded as actual divergence from the reference sequence.

2.11 Allele exchange crosses

rpoD is an essential gene, and it is the middle gene in a three gene operon. To avoid possible disruption of the *dnaG-rpoD-spr0980* operon, candidate suppressor alleles were transferred directly by transformation, without any linked selective marker gene, by taking advantage of the high replacement frequency achieved during natural genetic transformation with large, pure donor gene fragments. To replace one *rpoD* allele with another, a 5.5-kb amplicon encompassing *rpoD* and flanking genes, prepared by PCR from cells from a donor strain template with primers YT30 and YT31, as described above, was used as a donor at 100 ng/ml to transform a competent culture of the recipient strain for 80 minutes. After plating, individual colonies were isolated and resuspended in 10 ml of CAT+GP, outgrown until OD_{550} 0.2, and frozen in 10% glycerol. These isolates were scanned for the replacement allele by SNP determination as described in section 2.12. For the WT recipient, the frequency of allele replacement under these conditions averaged 60%. Approximately 1 out of 10 colony samples yielded a mixed result, indicating incomplete segregation of the transformed allele. Mixed colonies were re-streaked, and subclones were again tested for *rpoD* sequence as described in section 2.12. A purified subclone of each positive colony, verified by sequencing the entire *rpoD* gene, was retained for further analysis. For the *rpoD* mutant strains receiving the WT copy of *rpoD*, the frequency of allele replacement under these conditions averaged ~2%. To increase the frequency of identified transformants in these strains, 5 individual colonies were pooled and the pool was scanned for replacement alleles by SNP determination as described in section 2.12. Positive pools were re-streaked, and subclones were tested for the WT *rpoD* sequence as described in section 2.12. A

positive subclone was outgrown until OD_{550} 0.2 and frozen in 10% glycerol, purified subclones were verified by sequencing the entire *rpoD* locus, and retained for further analysis.

2.12 SNP determination by mismatched-primer assay

rpoD alleles were distinguished by a mismatched-primer PCR assay. Primers were designed to match ~20 bases, including each SNP as the 3′-terminal base of the primer: YT41, YT44, YT47, and YT50. Identical primers, except for the 3′-terminal base, which matched the WT sequence, were also created: YT40, YT43, YT46, and YT49. They were paired with a second primer, YT42, YT45, YT48, or YT51, to amplify a 500-bp diagnostic fragment. Thus, primer set YT40, YT41, and YT42 was used to detect the base change corresponding to the A171V amino acid substitution. Using YT40 or YT41 as the forward primer and YT42 is the reverse primer, two separate reactions were performed. The two reaction products were compared on a gel for the presence or absence of a band. The same was done with primer sets YT43, YT44, and YT45 for R316H; YT46, YT47, and YT48 for R355H; and YT49, YT50, and YT51 for L363F. Detection of a product with the WT primers indicated a WT *rpoD* sequence, while a product with only the primers matching the mutant allele indicated a mutant *rpoD* allele. Amplification was performed in 10-µl reaction mixtures, with 0.2 µl Phire HotStart II polymerase (Thermo Scientific) and Phire Reaction Buffer, 1 ng template DNA, 200 µmol/liter of each dNTP, and 0.5 µmol/liter of each primer. The amplification conditions were 98°C for 90 s and then 98°C for 15 s, 69°C for 15 s, and 72°C for 15 s for 20 cycles, followed by a 5-min 72°C final extension.

2.13 β-galactosidase assay

To measure β-galactosidase activity, a culture was induced to competence at an OD_{550} of 0.1 by the addition of 0.1 µg/mL DNA, 250 ng/mL CSP, 0.5 mM $CaCl_2$, and 0.04% BSA. At 10 minute intervals, 0.5-mL samples were removed, mixed with 0.125 mL lysis buffer (300 mM

Na_2HPO_4, 200 mM NaH_2PO_4, 50 mM KCl, 5 mM $MgSO_4$, 0.5% Triton X-100, and 250 mM β-mercaptoethanol), held at 37°C for 10 minutes, then put on ice. 0.1 mL of each mixture was loaded into a well in a 96-well plate, in duplicate. 0.05 mL of o-nitrophenyl-β-D-galactopyranoside solution was added to each well (4 mg/mL o-nitrophenyl-β-D-galactopyranoside, 60 mM Na_2HPO_4, 40 mM NaH_2PO_4). The plate was incubated at 37°C with shaking, and absorbance at 420 nm was read every 10 minutes for 100 minutes on the BioTek Synergy 2. The slope of the absorbance curve was used to calculate LacZ activity, which is reported in Miller units.

2.14 Competence kinetics assays.

To measure the amount of transformants at several time points after CSP addition, transforming cells were incubated with DNA for 5 minutes and then exposed to DNAseI. DNaseI was prepared as a 5 mg/mL solution in 0.15 M NaCl, and stored at -20°C. A log phase culture at OD_{550} 0.05 was treated with 250 ng/mL CSP, 0.5 mM $CaCl_2$, and 0.04% BSA and incubated at 37°C. Every 10 minutes post CSP addition, 0.1 ng/mL DNA was added to 1 mL cells and incubated at 37˚C for 5 minutes. DNAseI was added to a final concentration of 20 µg/mL, and the cells were incubated for another total of 75 minutes. Cells were chilled on ice for 5 minutes and then diluted and plated using the method described above.

2.15 Competent subpopulations assay.

To measure competent subpopulations, a log phase culture at OD_{550} 0.05 was treated with 250 ng/mL CSP, 0.5 mM $CaCl_2$, and 0.04% bovine serum albumin and incubated at 37°C, and 50 ng each of PCR amplified Em^R and Nv^R donor DNA. After 80 minutes of DNA exposure, cells were chilled for 5 minutes on ice and plated as described above on single and double drug plates.

2.16 Construction of chimeric *S. mutans/S. pneumoniae* σ^A, strain CP2464

To construct a *S. pneumoniae* strain carrying a chimeric σ^A, the upstream genes and region 1 of *S. pneumoniae rpoD* were amplified using primers YT66 and YT67. The downstream flank of *S. pneumoniae rpoD* was amplified using primers YT68 and YT69. YT67 and YT68 carried a BtsI restrictive enzyme recognition site. Regions 2-4 of *S. mutans rpoD* was amplified using primers YT104 and YT105, both with BtsI sites that, when cut, were complimentary to ends from primers YT67 and YT68, respectively. Purified amplification products were digested in 50-µL reactions with 10 units of BtsI (New England Biolabs), 1X NEBuffer 4 (New England Biolabs), 1 µg of DNA, up to 50 µL with dH_2O, for 1 hour at 55˚C. After one hour, digestion products were purified using a Clean and Concentrate kit (Zymo Research). 1.2 µg of the upstream and downstream fragments from *S. pneumoniae* were ligated to 400 ng of the *S. mutans rpoD* fragment, in a 20-µL reaction with 1 unit of T4 Ligase (New England Biolabs), 1X T4 DNA Ligase Buffer (New England Biolabs), and dH_2O to 20 µL, for 1 hour at room temperature. Ligase was then inactivated at 65˚C for 10 minutes. The products of ligation were transformed into competent *S. pneumoniae* cells as described above. To test individual clones for transformation, individual colonies were tested by mismatch PCR as previously described (Tovpeko & Morrison 2014), and in section 2.12. One positive clone was re-streaked, and one positive subclone was verified by Sanger sequencing, and retained as strain CP2464. The primers used to determine transformants were YT106 and YT107, while the primers used to determine non-transformants were YT107 and YT108.

2.17 Immunoblotting

To measure the amount of σ^X present during competence in the WT and the Δ*comW* mutant, strains CP2137 and CP2463 were grown in CAT+GP to an OD_{550} of 0.1 at 37°C. Cells were

induced to competence with 250 ng/mL CSP, 0.5 mM $CaCl_2$, and 0.04% BSA. 10-mL samples were collected every 5 minutes after induction and chilled rapidly, but not frozen, on dry ice. The samples were stored at 4°C until all samples were collected, centrifuged at 4500 rcf in a swinging bucket rotor at 4°C, resuspended in 0.7 mL of PBS and spun again in 1.5 mL microcentrifuge tubes. The pellet was resuspended in 0.3 mL of lysis buffer (10 mM Tris-HCl pH 8.0, 1 mM-EDTA, 1 mM-PMSF, 0.4% Triton X-100, 10 μg/mL DNAseI, 5 mM MgC1, and 5 mM CaCl2, and incubated at 37°C for 10 minutes. After adding Sarkosyl to 0.05% and EDTA to 5 mM, incubation continued at 37°C for another 3 minutes. Then trichloroacetic acid (TCA) was added to 10% to precipitate the protein. The samples were centrifuged at 10,000 rcf at 4°C for 10 minutes. The TCA supernatant was removed and the pellet was washed twice by adding 1 mL of acetone and spinning at 10,000 rcf at 4°C for 10 minutes. Then the pellet was dried in the hood until the acetone had evaporated. The sample was resuspended in 40 μL of PBS and 10 μL of loading buffer (50mM Tris-HCl, pH 6.8, 110 mM β-mercaptoethanol, 2% sodium dodecyl sulfate (SDS), 0.1% bromophenol blue, and 10% glycerol). Samples were heated at 95°C for 10 minutes and subjected to SDS gel electrophoresis using 4-20% Tris-glycine gels (Bio-Rad Laboratories, Hercules, CA). Electrophoresis at 180V in Tris-glycine buffer was followed by transfer to a 0.2 μm polyvinylidene diflouride membrane (Bio-Rad Laboratories, Hercules, CA) for 2 hours at 36 V in Transfer Buffer (48 mM Tris, 39 mM glycine, pH 9.2). The membrane was blocked for 1 hour at room temperature in TBST buffer (20 mM Tris-HCl, pH 7.6, 140 mM NaCl, and 0.1% Tween-20) containing 5% nonfat milk (Difco). The membrane was the probed with purified polyclonal antibody against σ^X (custom rabbit antibody, GenScript, Piscataway, NJ) and purified mouse monoclonal antibody against *E.* coli-σ^{70} (Biolegend, #663202, San Diego, CA) at dilutions of 1:1000 and 1:5000, respectively, in TBST with 5% milk at 4°C overnight. After being immersed in 25 mL of TBST

for 5 minutes three times at room temperature, the membrane was incubated with secondary peroxidase-conjugated goat anti-rabbit and goat anti-mouse immunoglobulin G antibodies at dilutions of 1:10,000 (Sigma-Aldrich, St. Louis, MO) for 1 hour at room temperature in TBST containing 5% milk then washed with 25 mL of TBST three times. The signal was detected using SuperSignal West Dura Extended Duration Substrate (ThermoFisher, Waltham, MA) on the Chemiluminescence channel of the Odyssey FC (LI-COR Biosciences, Lincoln, NE). The band intensities were measured using AlphaImager (ProteinSimple, San Jose, CA).

III. Competence for Genetic Transformation in *Streptococcus pneumoniae*: Mutations in σ^A Bypass the *comW* Requirement.

(Portions of this chapter were published in: Journal of Bacteriology. 196: 3724-3734, 2014)

3.1 Abstract

Competence for genetic transformation in the genus *Streptococcus* depends on an alternative sigma factor, σ^X, for coordinated synthesis of 23 proteins, which together establish the X state by permitting lysis of incompetent streptococci, uptake of DNA fragments, and integration of strands of that DNA into the resident genome. Initiation of transient accumulation of high levels of σ^X is coordinated between cells by transcription factors linked to peptide pheromone signals. In *Streptococcus pneumoniae*, elevated σ^X is insufficient for development of full competence without coexpression of a second competence-specific protein, ComW. ComW, shared by eight species in the *Streptococcus mitis* and *Streptococcus anginosus* groups, is regulated by the same pheromone circuit that controls σ^X, but its role in expression of the σ^X regulon is unknown. Using the strong, but not absolute, dependence of transformation on *comW* as a selective tool, we collected 27 independent *comW* bypass mutations and mapped them to 10 single-base transitions, all within *rpoD*, encoding the primary sigma factor subunit of RNA polymerase, σ^A. Eight mapped to sites in *rpoD* region 4 that are implicated in interaction with the core β subunit, indicating that ComW may act to facilitate competition of the alternative sigma factor σ^X for access to core polymerase.

3.2 Introduction

Streptococcus pneumoniae is a common human pathogen, usually carried asymptomatically in the nasopharynx. A prominent characteristic of the bacterium is a natural ability to take up exogenous DNA from its environment. If the exogenous DNA is integrated into the genome, the bacterium is said to be transformed. Such transformation depends on development of a specialized physiological state, termed competence, development of which is coordinated within a culture by a quorum-sensing (QS) mechanism encoded by two genetic loci, *comAB (Pestova et al. 1996)* and *comCDE (Hui & Morrison 1991)*. Both loci are transcribed at a basal level by the "housekeeping" sigma factor, σ^A (Morrison & Jaurin 1990; Martin et al. 2013). The competence-stimulating peptide (CSP), a product of the *comC* gene, is secreted by an ABC transporter/protease encoded by *comA* and *comB (Martin et al. 2013; Lacks 2004)*. CSP is sensed by a histidine kinase receptor, ComD, which phosphorylates a cognate response regulator, ComE (Pestova et al. 1996; Martin et al. 2013; Håvarstein et al. 1996; Ween et al. 1999). Phosphorylated ComE activates the promoters of eight operons comprising 13 genes transcribed specifically at competence by binding a direct repeat centered at −40, the ComE box (Martin et al. 2013; Oggioni & Morrison 2008; Martin et al. 2010). They are designated early competence genes (Martin et al. 2010) and include both the *comAB* and *comCDE* operons (Peterson et al. 2004; Peterson et al. 2000; Havarstein et al. 1995; Lee & Morrison 1999). This organization creates a positive-feedback loop, ensuring a rapid increase in the level of CSP that can cause all the cells in a culture to become competent simultaneously (Ween et al. 1999; Lee & Morrison 1999). One additional early gene essential for competence is *comX*, encoding the alternative sigma factor σ^X, which accumulates to high levels in response to CSP (Luo et al. 2003), forming a holoenzyme with RNA polymerase (RNAP) and enabling recognition of a noncanonical promoter sequence termed the ComX box, or

cinbox, upstream of 23 genes, termed the late competence genes (Lee & Morrison 1999; Luo et al. 2003; Luo & Morrison 2003; Morrison 2007).

Streptococci have only one known alternative sigma factor, which in all species is linked to expression of the late competence genes. In the *Streptococcus mitis* and *Streptococcus anginosus* groups of species, upstream regulators of *comX* include CSP/ComABCDE QS systems orthologous to the pneumococcal one (Håvarstein 2010). In all other groups of streptococci, *sigX* is regulated by a different class of peptide QS circuit, using genes named *comS* and *comR*, encoding a peptide pheromone and intracellular receptor, respectively (Håvarstein 2010). Regulation of σ^X activity is important, as competence genes encode one or more killer proteins and cognate immunity genes. Within the *S. mitis* and *S. anginosus* groups, eight related species (*S. anginosus*, *Streptococcus cristatus*, *Streptococcus infantis*, *S. mitis*, *Streptococcus oligofermentans*, *Streptococcus oralis*, *S. pneumoniae*, and *Streptococcus pseudopneumoniae*) add another layer of control of σ^X activity in the form of a small protein, ComW, which plays an important role during competence development. Since its identification in 2004 in *S. pneumoniae* (Luo et al. 2004), attempts to identify the role of ComW have been reported (Luo et al. 2004; Sung & Morrison 2005; Piotrowski et al. 2009; Weng et al. 2013), but its function remains unclear. *comW* mutants have several phenotypes that suggest different possible roles for the protein. First, *comW* mutants are ~10^4-fold deficient in transformants (Luo et al. 2004). Second, in *comW* mutants, transcription of late competence genes is strongly reduced, to ~10% of wild-type (WT) levels, and accumulation of σ^X protein is similarly reduced (Sung & Morrison 2005). Third, *comW clpE* (where ClpEP is a protease that degrades σ^X) double mutants produce restored amounts of σ^X protein but are still deficient in transformation (Luo et al. 2004). Fourth, both the N and C termini of ComW are necessary for σ^X stability, while the N terminus is necessary for

σ^X function (Sung & Morrison 2005). Since none of the roles of ComW are well characterized and available data suggest several alternative roles during transformation, ComW's function in supporting σ^X activity remains unknown.

To investigate the role of ComW in regulation of competence in an unbiased manner, we designed and carried out a suppressor screen for rare mutations that could bypass the *comW* requirement and restore the competence lost in Δ*comW* strains. Here, we describe a strategy for recovering rare *comW* bypass mutations and show that they map exclusively to *rpoD*.

3.3 Results

3.3.1 Rare Δ*comW* suppressors are recovered by serial enrichment.

The expression of genes regulated by an alternative sigma factor typically depends on numerous additional factors affecting the entire life cycle of the sigma factor itself, including its synthesis, its availability, its stability, its success in competing for occupancy of limiting core polymerase, and interaction of the resulting holoenzyme with specific promoter sites. As existing data indicate that ComW might affect various levels in the life cycle of σ^X, including synthesis, stability, and activity, but no molecular mechanism has yet been identified, we wished to obtain an unbiased genetic indication of the most important site of action of ComW by collecting and mapping suppressor mutations that could partially bypass the requirement for *comW* in competence development.

Although the severe transformation deficiency of *comW* mutants (10,000-fold reduced) offers the possibility of strong enrichment for such suppressors in a single step of transformation, a preliminary screen of a transposon library (provided by P. Burghout, Nijmegen, The Netherlands) did not yield any such suppressors; because transposon insertions produce predominantly null mutations, this may indicate that important players might be essential for viability (data not

shown). To allow recovery of suppressors that were not null mutations, we sought conditions that would provide a sufficiently rich array of base substitutions to allow recovery of more subtle bypass mutations but not such a high density of base changes as to make mapping by WGS cumbersome. Because strain Rx1 carries a mutation inactivating the HexAB mismatch repair system, which creates a mild mutator phenotype (Tiraby & Fox 1973), we conducted a pilot enrichment in the Rx1 derivative CP2463, in which *comW* was replaced by a Kn^r cassette and a *comA* deletion made development of competence absolutely dependent on exogenous CSP.

A single *comW* mutant colony of strain CP2463 was grown for ~35 doublings to create a library of mixed, potentially mutant subclones (Fig. 9). The total frequency of trimethoprim-resistant mutants in a similar culture was estimated as 0.01% (data not shown), indicating the presence of numerous spontaneous base pair changes in the pool. Subjecting this library to transformation could preferentially retrieve mutations that suppress the loss of transformation caused by ComW deficiency, but as the transformation efficiency of a Δ*comW* mutant is nonzero (10^{-4}), transformants of rare-suppressor-bearing mutants would not be expected to dominate the pool of transformants obtained from a single round of transformation. However, we anticipated that the enrichment achieved in this pool would be compounded by a second round of transformation of the entire transformant pool and that suppressor mutants might predominate after further repetition of the enrichment.

To maximize recovery of rare suppressors in the first cycle of enrichment, a pure 7.4-kb Nv^r DNA amplicon, which routinely transforms the WT with 80 to 90% efficiency, was chosen to transform a portion of the mutant library (100 million cells). Transformant colonies (2×10^4) were collected as a pool of Nv^r cells and outgrown to produce a working pool stock. This pool was not rich in suppressors (Table V); indeed, the recovered Nv^r library did not transform

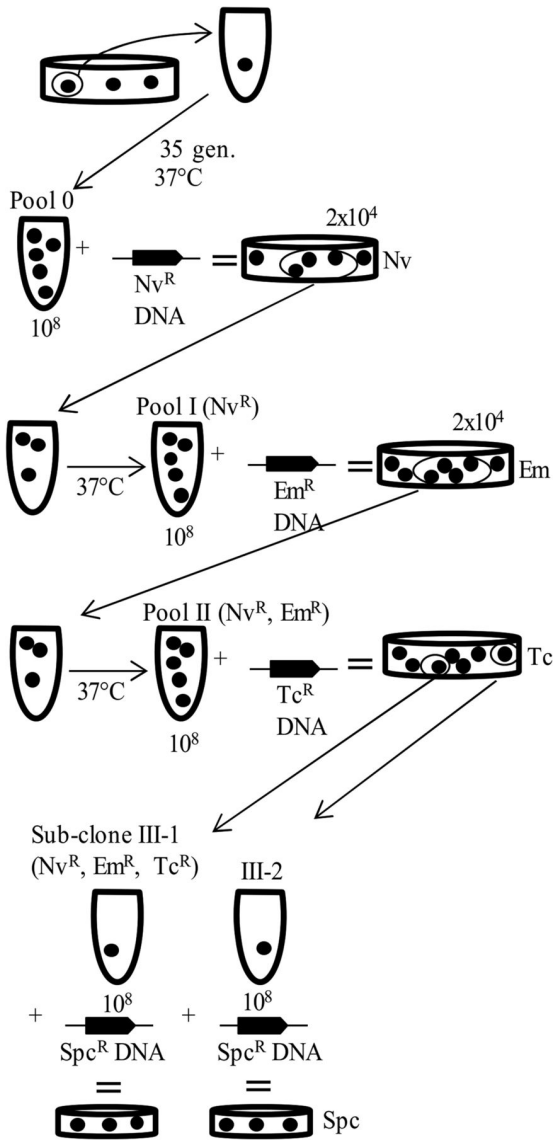


FIGURE 9. Strategy for enrichment of *comW* bypass mutants. A single colony was picked and grown for 35 generations to create pool 0, a library of potential suppressor mutations. Then 10^8 cells of pool 0 were transformed with Nv^r DNA and plated, and 2 X 10^4 Nv^r transformant colonies were collected to create pool I. Next, 10^8 cells of pool I were transformed with Em^r DNA, and 2 X 10^4 Nv^r Em^r transformant colonies were collected to create pool II. After 10^8 cells of pool II were transformed with Tc^r DNA, individual Nv^r Em^r Tc^r transformant colonies were picked and transformed with Spc^r DNA to determine the transformation efficiency. The dots in tubes represent cells; the dots on plates represent colonies.

detectably more than the Δ*comW* background rate of 10^{-4} compared to the WT strain. The pool of Nv^r transformants was next transformed with a genomic Em^r donor DNA. Among 20 resulting Em^r transformants, none transformed better than the Δ*comW* mutant (data not shown). Since suppressors were thus clearly not dominant among the resulting Nv^r Em^r transformants, a second enrichment library was collected as a pool of 20,000 of the Nv^r Em^r transformants (Fig. 9) and subjected to a third cycle of transformation using genomic Tc^r donor DNA (Fig. 9). Among six resulting Nv^r Em^r Tc^r transformants, four transformed at a much higher frequency (400- to 2,000-fold) than the *comW* mutant parent (Table V). This indicates that suppressors with increased transformation efficiencies dominated the resulting Nv^r Em^r Tc^r transformant collection and established that three cycles of enrichment by this procedure could suffice to recover rare stable *comW* bypass mutants.

To obtain additional independent bypass mutants, this compound enrichment process was repeated in parallel for 13 additional separate single-colony subclones of CP2463, following the same series of successive enrichment cycles. After the first enrichments, 9 of the 14 Nv^r pools transformed at slightly elevated rates, while two transformed 10 times as well as the *comW* parent, indicating that suppressor mutants were present in the pools, but at a low frequency (Fig. 10 and Table VI). After the second series of enrichment cycles, 5 of the 14 Nv^r Em^r pools transformed at rates at least 100-fold above that of the *comW* parent (Table VI). After the third enrichment, among five Nv^r Em^r Tc^r transformant subclones tested from each enrichment, at least one transformed at a rate at least 20-fold higher than that of the Δ*comW* parent (Fig. 10 and Table VII), indicating that suppressor mutants dominated in all 14 cases.

We conclude that rare, spontaneously arising suppressors of the c*omW* transformation deficiency can be recovered repeatedly by three cycles of compounded enrichments.

TABLE V: Recovery of suppressor mutants through serial transformational enrichment.

Culture[a]	Transformation Efficiency[b]		Relative Efficiency[e]
	WT[c]	Mutant[d]	
Spontaneous Mutation Library (Pool 0)	$5x10^7$	$2x10^4$	0.0004
Collected Nv^R Library (Pool I)	$5x10^5$	$1x10^2$	0.0002
Collected Nv^REm^R Library (Pool II)	$5x10^5$	$5x10^2$	0.001
$Nv^REm^RTc^R$ Clone 1	$1x10^5$	$3x10^2$	0.003
$Nv^REm^RTc^R$ Clone 2	$1x10^5$	$9x10^1$	0.0009
$Nv^REm^RTc^R$ Clone 3	$1x10^5$	$4x10^3$	0.04
$Nv^REm^RTc^R$ Clone 4	$1x10^5$	$1x10^4$	0.1
$Nv^REm^RTc^R$ Clone 5	$1x10^5$	$1x10^4$	0.1
$Nv^REm^RTc^R$ Clone 6	$1x10^5$	$2x10^4$	0.2

a. Pool or sub-clone in enrichment series PRT.

b. Transformation efficiency: drug-resistant cfu/mL, determined in quadruplicate. PCR marker used for Pool 0, genomic markers used for Pools I, II, and clones 1-6. SD values below 20%.

c. WT control

d. Mutant pool or clone

e. Transformation efficiency of strain or pool relative to WT

TABLE VI: Relative transformation efficiency of parallel pools during enrichments.

Independent Enrichment[a]	Relative Efficiency[b]		
	Pool 0	Pool I	Pool II
PR	0.00038	0.00022	0.00113
IL	0.00011	0.00047	0.00903
NY	0.00016	0.00039	0.00800
AL	0.00009	0.00002	0.00627
DE	0.00009	0.00009	0.00366
FL	0.00006	0.00100	0.02032
GA	0.00004	0.00003	0.01094
MA	0.00004	0.00041	0.00359
MT	0.00004	0.00038	0.00509
NV	0.00010	0.00076	0.01219
OH	0.00003	0.00035	0.00063
NC	0.00003	0.00022	0.00229
SC	0.00008	0.00107	0.01232
TX	0.00008	0.00212	0.01414
WT	1	1	1
ΔcomW	0.0001	0.0001	0.0001

a. Name of enrichment series

b. Relative transformation efficiency, as drug resistant cfu/mL/WT. SD values of triplicate measurements were below 20%

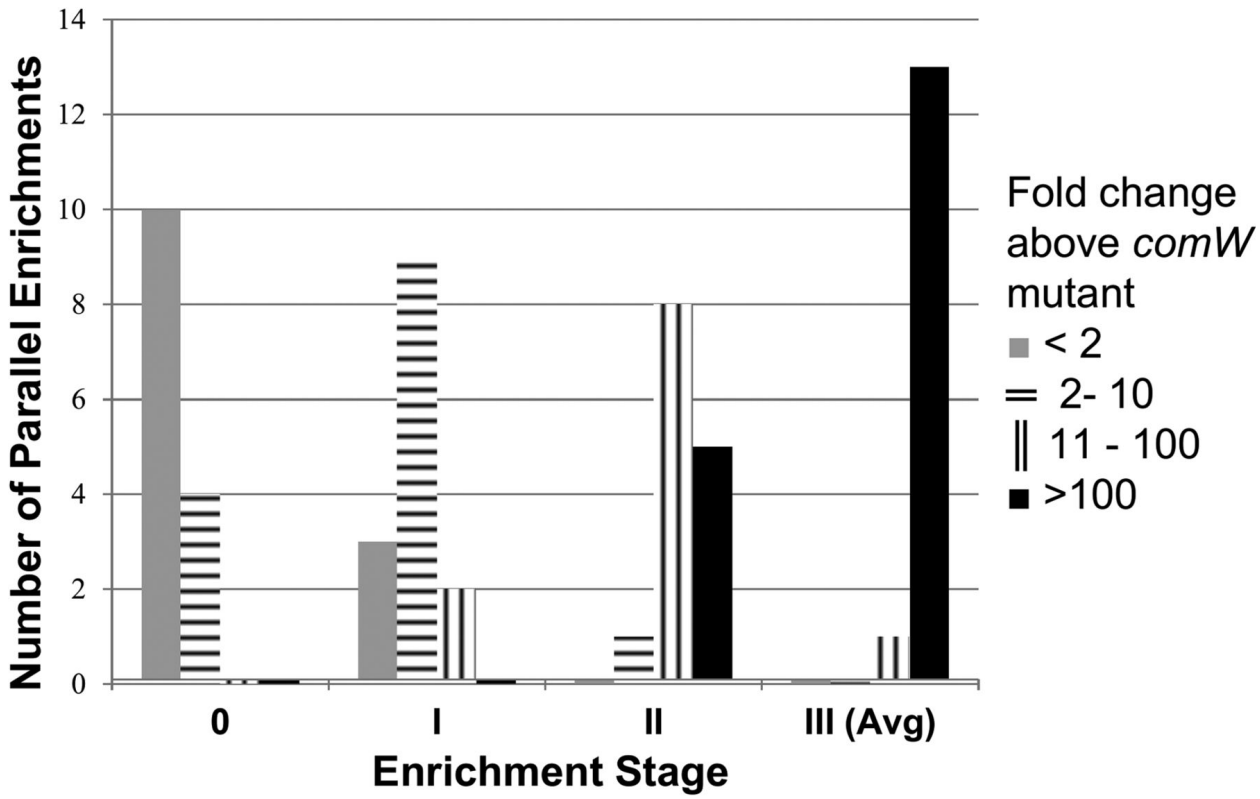


FIGURE 10. Serial enrichment of suppressor mutants from 14 independent libraries. Shown are counts of pools with transformation efficiencies in four ranges, for the initial mutant library (stage 0) and 3 successive cycles of enrichment. Values for enrichment stage III reflect average (Avg) of subclone efficiencies (Table VII).

TABLE VII: σ^A substitutions in bypass mutants.

Independent enrichment[a]	Individual[b]	Relative Efficiency[c]	σ^A Substitution
ALT	5	0.002	R297C
	2	0.007	R316C
	1	0.01	P287S
	3	0.02	A171V
	4	0.05	A171V*
DET	1	0.0003	-
	3	0.006	A171V
	4	0.009	-
	5	0.05	A171V*
FLT	3	0.004	R355H
	5	0.006	R355H
	1	0.03	R316H
	4	0.08	R355H*
GAT	5	0.00008	-
	4	0.02	R316H
	2	0.06	R316H
	3	0.08	A171V
	1	0.09	R316H*
ILT	3	0.0004	-
	4	0.1	R316H
	2	0.26	R316H
	1	0.3	R316H
	5	0.31	R316H*
MAT	2	0.002	-
	1	0.003	-
	4	0.003	Y290C
	3	0.01	R316H
	5	0.02	A171V*

a. Name of enrichment series

b. Tc^R clone number

c. Relative transformation efficiency, as drug-resistant cfu/mL/WT. SD values of quadruplicate measurements were below 20%.

*, mutation identified by WGS; -, isolate not sequenced.

TABLE VII (continued)

Independent enrichment[a]	Individual[b]	Relative Efficiency[c]	σ^A Substitution
MTT	4	0.008	R316H
	3	0.03	A171T
	2	0.06	A171V
	1	0.07	R316H
	5	0.07	R316H*
NCT	1	0.001	-
	3	0.02	A171V
	5	0.02	A171V
	2	0.07	A171V*
	4	0.07	A171V
NVT	2	0.003	V314A
	4	0.02	A171V
	5	0.02	A171V
	3	0.07	V314A
NYT	1	0.1	R316H*
	5	0.00008	-
	4	0.008	L363F
	1	0.18	L363F*
	3	0.28	R316H
	2	0.41	R316H*
OHT	1	0.05	R355H*

a. Name of enrichment series

b. Tc^R clone number

c. Relative transformation efficiency, as drug-resistant cfu/mL/WT. SD values of quadruplicate measurements were below 20%.

*, mutation identified by WGS; -, isolate not sequenced.

TABLE VII (continued)

Independent enrichment[a]	Individual[b]	Relative Efficiency[c]	σ^A Substitution
PRT	2	0.0009	-
	1	0.003	-
	3	0.04	R355H
	4	0.1	R316H
	5	0.1	R316H
	6	0.2	R316H
SCT	5	0.08	R316H
	4	0.1	R316H*
	2	0.13	R316H
TXT	2	0.06	R316H
	5	0.1	R316H
	4	0.11	R316H*
WT		1	WT
ΔcomW		0.0001	WT

a. Name of enrichment series

b. Tc^R clone number

c. Relative transformation efficiency, as drug-resistant cfu/mL/WT. SD values of quadruplicate measurements were below 20%.

*, mutation identified by WGS; -, isolate not sequenced.

3.3.2 All suppressor mutants contain a SNP in *rpoD*.

Since competence was substantially restored in each of at least 14 independently derived Δ*comW* mutants, we sought to learn whether suppression was caused by single-gene mutations and to identify mutations that might be linked to the suppressor phenotype. To identify candidate unmarked mutations, WGS was carried out on a total of 16 strains, including 14 suppressor mutants, the Δ*comW* parent (CP2463), and the corresponding $comW^+$ wild type (CP2137). The sequencing reads were aligned to an annotated reference sequence of *S. pneumoniae* strain R6 (Hoskins et al. 2001), which is derived, like Rx1, from the clinical isolate D39 (Tiraby et al. 1975). Approximately 9,000 SNPs were identified in total (~600 per sequenced strain) (Table VIII and Appendix 1). After removing from consideration SNPs in regions with less than 40-fold coverage, ~6,600 SNPs mapped to coding regions as identified in the R6 genome annotation. Of these, ~2,700 cause synonymous amino acid changes. Among the remaining 3,900 SNPs, 3,768 reflected divergence of the Rx1 and R6 lineages (Tiraby & Fox 1973), i.e., differed from the R6 sequence but were present in all of our sequenced strains (14 suppressors, the WT, and the Δ*comW* parent). Among the remaining 132 SNPs, 64 were mapped to donor drug markers introduced during enrichment cycles. The remaining 68 SNPs of interest were distributed among 53 unique genes among the sequenced suppressor strains, ranging from 2 to 10 SNPs per strain. However, only the gene *rpoD* contained a SNP in each of the 14 sequenced suppressor strains (Table VIII and Table S1 in Appendix). As the remaining 54 SNPs were scattered among 52 other genes (Fig. 11A), we hypothesized that the single gene hit in every suppressor strain, *rpoD*, was the site of the effective bypass mutations.

TABLE VIII: Filtering of SNPs detected by WGS in 14 suppressor strains.

Stage of Analysis	Total SNPs in 14 strains	SNPs Per Strain
All SNPs detected vs. R6[a]	9000	640
High quality seq. coverage[b]	7500	540
Changes in ORFs[c]	6600	470
Non-synonymous	3900	280
Absent from parent[d]	132	9
Outside introduced markers[e]	68	5
In *rpoD*	14	1

a. Total SNPs after alignment to accession NC_003098.1
b. SNPs remaining after removal of SNPs with poor sequence coverage
c. ORFs as annotated in NC_003098.1
d. SNPs absent from Δ*comW* parent
e. SNPs not introduced by Nv^R, Em^R, or Tc^R transformation

Using Sanger sequencing of PCR products, we confirmed the presence of each *rpoD* mutation in the 14 suppressor strains used for WGS and the absence of any *rpoD* mutation from the WT, the Δ*comW* parent, and the Nv^r, Tc^r, Spc^r, and Em^r donor strains, indicating that the mutations arose during the mutagenesis and enrichment cycles. When *rpoD* was similarly sequenced in the remaining 37 suppressor mutants identified among cycle 3 Nv^r Em^r Tc^r transformants (Table VII), several more independent instances of the four *rpoD* mutations identified by WGS were revealed, as well as six additional *rpoD* nucleotide changes, as shown in Fig. 11B.

Altogether, the mapping and sequence data supported an inference that bypass of the *comW* requirement is available solely or principally by specific modification of σ^A.

3.3.3 Four mutant *rpoD* alleles substantially rescue transformation in Δ*comW* strains and account for most or all of the suppression phenotype.

To determine directly whether the mutations identified in *rpoD* could bypass the loss of competence in *comW* mutants, two sets of allele replacement crosses were done for four representative *rpoD* suppressor alleles, including the three that had been recovered most frequently. Each selected *rpoD* allele was transferred to a WT background, followed by deletion of *comW* to allow detection of possible suppression. Conversely, the *rpoD** allele (where the asterisk indicates any of the four mutant *rpoD* alleles) in each of the four corresponding suppressor mutants was replaced by the wild-type sequence (Fig. 12A). Since the mutations were not linked to a selective marker, we used a single-nucleotide primer mismatched PCR assay (Fig. 12C) to distinguish *rpoD* transformants from untransformed progeny. In all four cases, the backcross of mutant *rpoD* alleles was sufficient to increase transformation in a *comW* background to the level

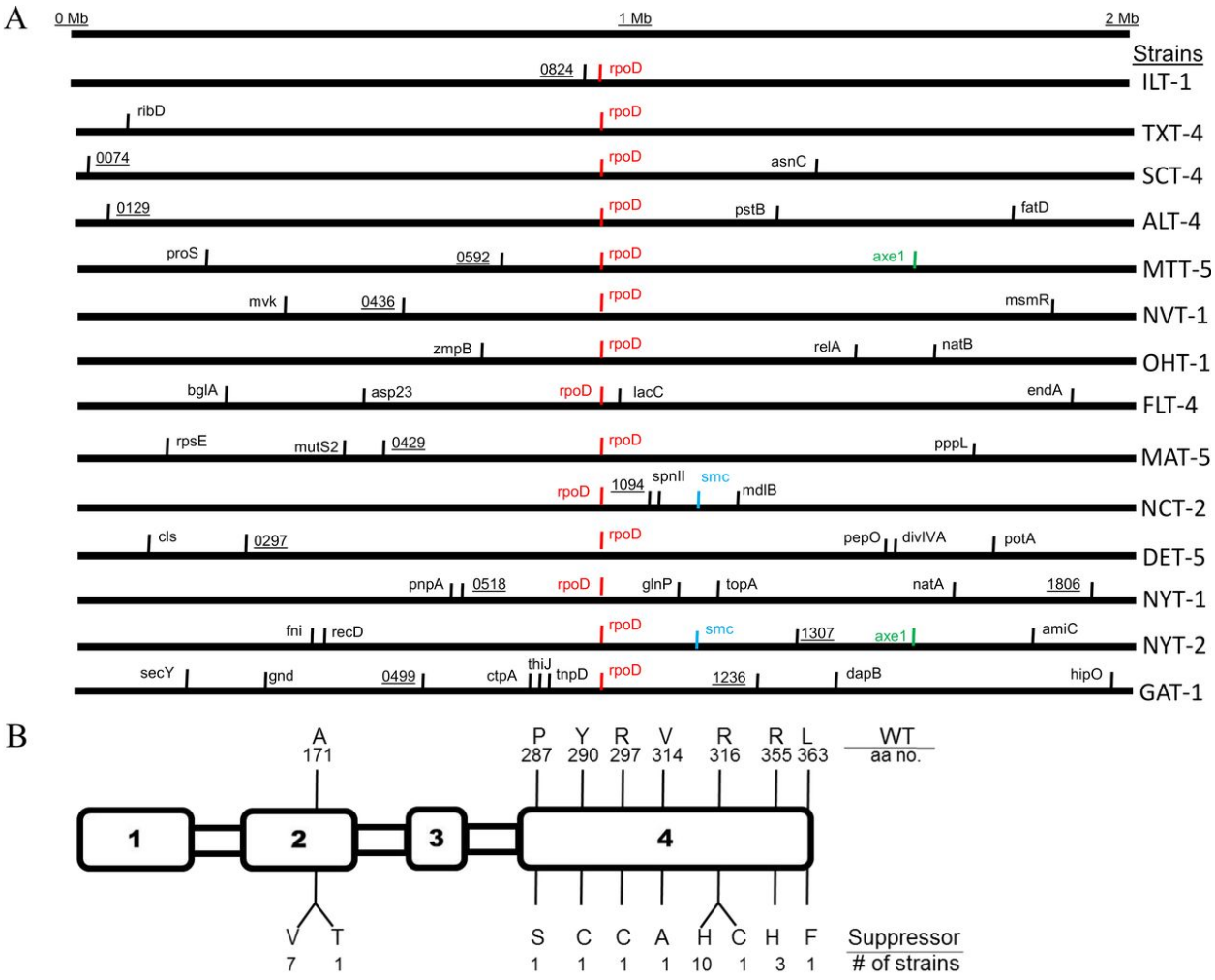
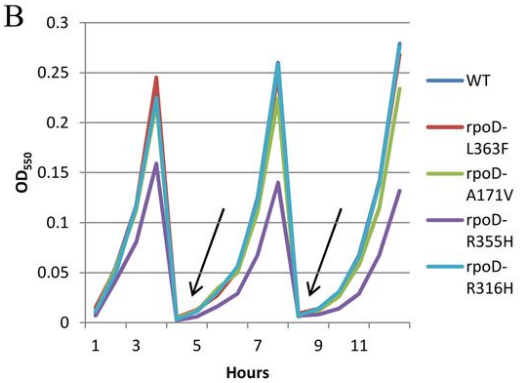


FIGURE 11. Locations of 68 amino acid changes identified by whole-genome and *rpoD*-targeted sequencing. (A) Mapping of base changes identified by WGS in 14 independent *comW* bypass mutants. Non-synonymous substitutions are organized by relative genome positions, and gene designations are as annotated in accession number NC_003098.1. Black, unique gene hit; color, 2 or more hits in the same gene; black horizontal line, entire *S. pneumoniae* genome. The names of the bypass strains sequenced by WGS are on the right. (B) Map of predicted amino acid residue changes in RpoD (σ^A) among 27 suppressor mutants (Table VII). Boxes, four conserved regions of σ^A, 369 amino acids (aa), as assigned by Vassylyev et al. (Vassylyev et al. 2002). The WT residues and positions are shown above the protein, and the suppressor residues and numbers of cases are shown below.

A

Strain	Description	comW	rpoD	Relative Transformation Efficiency (%) A171V	R316H	R355H	L363F
1	WT	■	●	100	100	100	100
2	Δ*comW*	□	●	0.02	0.02	0.02	0.02
3	*rpoD** isolate	□	○	3	9	6	5
4	*rpoD** in WT	■	○	75	111	73	102
5	*rpoD** in WT, Δ*comW*	□	○	7	7	2	6
6	*rpoD*⁺ (cured) isolate	□	●	0.3	0.01	0.2	0.02

B



C

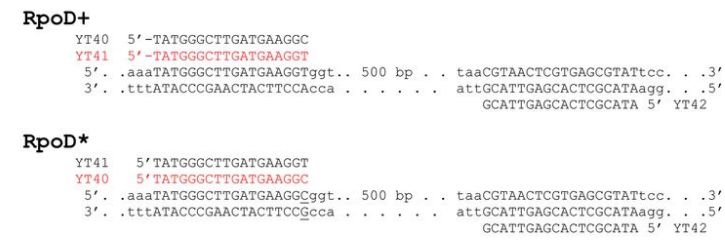


FIGURE 12. Linkage between *rpoD mutations and the *comW* bypass phenotype.** (A) Backcross analysis of the bypass phenotypes of four *rpoD** isolates. Shown is a comparison of the transformation efficiency of the suppressor isolate (line 3) to those of isolates with *rpoD** in a WT background (line 4) and with *rpoD** in a transformant Δ*comW* WT background (line 5) and a suppressor isolate cured of the *rpoD** mutation (line 6). Solid lines, WT genome; dashed lines, genome of isolate recovered from enrichment; ■, *comW*⁺; □, Δ*comW*; ●, *rpoD*⁺; ○, *rpoD** (the bypass mutant residues are indicated at the tops of the data columns. For mutations A171V, R316H, R355H, and L363F, the suppressor isolates used were ALT4, ILT1, FLT4, and NYT1, respectively; the strains were backcrossed *rpoD** alleles in a Δ*comW* WT background were CP2456, CP2458, CP2457, and CP2455; and the *rpoD*⁺ cured strains were CP2459, CP2461, CP2460, and CP2462. Standard deviation (SD) values were below 20%. (B) Growth of strains CP2451 to -54 containing *rpoD** mutations over 17 generations. The arrows indicate 1:100 dilutions of the exponentially growing culture in fresh medium. (C) Allele-specific alternative primer pairs used for mismatch PCR genotyping of segregants after transformations with 5.5-kb donor amplicons. The SNP in the A171V mutant is underlined. The productive allele-specific primer for each allele is shown in red.

observed in the original suppressor mutant (Fig. 12A). In each case, introduction of the putative suppressor allele in the WT background did not itself detectably affect the transformation efficiency or growth rate (Fig. 12A and B). However, after the removal of *comW*, each *rpoD** *comW* double mutant strongly suppressed the transformation deficiency typical of *comW* mutants. Thus, suppression of the *comW* transformation defect was indeed linked to the mutant *rpoD** alleles and not to any other mutations that may have accumulated during enrichment of suppressor mutants.

To test the linkage of suppression to the mutant alleles of *rpoD* further, we cured each of the four suppressor mutant types of its *rpoD** SNP by replacement with the WT *rpoD* (Fig. 12A). For each mutation, one suppressor isolate was selected at random for the cross, although three of the mutations arose more than once. On curing of their *rpoD* mutations, the four strains lost their suppression phenotype, two completely and two partially (Fig. 12A). In the former cases, this indicates that suppression is entirely explained by the mutations in *rpoD*. In the latter cases, it appears to be possible that an additional mutation is present in the recovered suppressor mutants that accounts for a modest level of suppression (~10-fold higher than that of the *comW* mutant). Nonetheless, as the corresponding backcross strains containing only the mutant allele of *rpoD* transformed at the level of the original suppressor mutant, it is clear that in both cases the mutant *rpoD* allele was primarily responsible for the *comW* bypass phenotype.

We conclude that each of the four substitutions in σ^A, A171V, R316H, R355H, and L363F, is individually both necessary and sufficient for *comW* bypass in the suppressor mutants recovered through repeated transformational enrichment.

3.4 Discussion

To seek clues to the critical function of ComW during competence development, rare suppressors were retrieved from a library of spontaneous mutants prepared in a *comW* mutant background. This strategy provided a minimally biased way to identify possible ComW interaction partners, especially by allowing the examination of essential genes. Whole-genome sequencing and linkage analysis identified 10 different spontaneous mutations that substantially increased transformation efficiency in the ComW-deficient background. Remarkably, all 10 suppressors mapped to *rpoD*, which encodes the essential primary sigma factor subunit of RNA polymerase. This immediately implicates the housekeeping sigma factor in the function of ComW during competence development in *S. pneumoniae* and may explain why a previous approach using a mariner transposon was not successful at identifying *comW* suppressors.

Although there is no currently available structure for the *S. pneumoniae* RNA polymerase holoenzyme, the strong conservation of the sequence and structure of bacterial RpoD proteins suggests it could be informative to inspect the structure of σ^{70} in the *Escherichia coli* holoenzyme for clues to the nature of the pneumococcal *comW* bypass mechanism. To identify homologous residues in *E. coli* and *Thermus aquaticus* σ^{70} that correspond to those affected in the pneumococcal bypass mutants, we aligned regions 2 to 4 of RpoD proteins from several species (Fig. 13). To make comparisons, we used *E. coli*, *T. aquaticus*, and *Staphylococcus aureus* sequences, for which there are available crystal structures, and the *Streptococcus mutans* sequence, because it is a *Streptococcus* species that is naturally transformable but does not have ComW. Region 4 of σ^{A}, the site of most of the bypass substitutions, forms a protrusion rooted by region 3 in the cleft between the β and β′ subunits (Fig. 14). In the shape of a slightly cupped left hand, it exposes several surfaces with known functions in the initiation of transcription. The base knuckles

```
T. aquaticus      AKKYTGRGLSFLDLIQEGNQGLIRAVEKFEYKRRFKFSTYATWWIRQAINRAIADQARTI     273
E. coli           AKKYTNRGLQFLDLIQEGNIGLMKAVDKFEYRRGYKFSTYATWWIRQAITRSIADQARTI     450
S. aureus         AKRYVGRGMLFLDLIQEGNMGLIKAVEKFDFNKGFKFSTYATWWIRQAITRAIADQARTI     206
S. pneumoniae     AKRYVGRGMQFLDLIQEGNMGLMKAVDKFDYSKGFKFSTYATWWIRQAITRAIADQARTI     206
S. mutans         AKRYVGRGMQFLDLIQEGNMGLMKAVDKFDYSKGFKFSTYATWWIRQAITRAIADQARTI     208
                  **:*. **: ********* **::**:**:: : :**************.*:********

T. aquaticus      RIPVHMVETINKLSRTARQLQQELGREPSYEEIAEAMGPGWDAKRVEETLKIAQEPVSLE     333
E. coli           RIPVHMIETINKLNRISRQMLQEMGREPTPEELAERMLMP--EDKIRKVLKIAKEPISME     508
S. aureus         RIPVHMVETINKLIRVQRQLLQDLGRDPAPEEIGEEMDLP--AEKVREILKIAQEPVSLE     264
S. pneumoniae     RIPVHMVETINKLVREQRNLLQELGQDPTPEQIAERMDMT--PDKVREILKIAQEPVSLE     264
S. mutans         RIPVHMVETINKLVREQRNLLQELGQDPTPEQIAERMDMT--PDKVREILKIAQEPVSLE     266
                  ******:****** *  *:: *::*::*: *::.* *      .::.: ****:**:*:*

T. aquaticus      TPIGDEKDSFYGDFIPDENLPSPVEAAAQSLLSEELEKALSKLSEREAMVLKLRKGLIDG     393
E. coli           TPIGDDEDSHLGDFIEDTTLELPLDSATTESLRAATHDVLAGLTAREAKVLRMRFGIDMN     568
S. aureus         TPIGEEDDSHLGDFIEDQEAQSPSDHAAYELLKEQLEDVLDTLTDREENVLRLRFGLDDG     324
S. pneumoniae     TPIGEEDDSHLGDFIEDEVIENPVDYTTRIVLREQLDEILDTLTDREENVLRLRFGLDDG     324
S. mutans         TPIGEEDDSHLGDFIEDEVIENPVDYTTRVVLREQLDEVLDTLTDREENVLRLRFGLDDG     326
                  ****::.**. **** *     * : ::   *    .. *  *: **  **::* *:

T. aquaticus      REHTLEEVGAYFGVTRERIRQIENKALRKLKYHESRTRKLRDFLE-      438
E. coli           TDYTLEEVGKQFDVTRERIRQIEAKALRKLR-HPSRSEVLRSFLDD      613
S. aureus         RTRTLEEVGKVFGVTRERIRQIEAKALRKLR-HPSRSKRLKDFMD-      368
S. pneumoniae     KMRTLEDVGKVFNVTRERIRQIEAKALRKLR-QPSRSKPLRDFIED      369
S. mutans         KMRTLEDVGKVFDVTRERIRQIEAKALRKLR-HPSRSKQLRDFVED      371
                     ***:**  * ********** ******: : **:. *:.*::
```

FIGURE 13. Alignment of σ^A homologs. *S. pneumoniae* σ^A amino acid residues 147 to 369 were aligned with σ^A sequences from accession numbers Q9EZJ8.1 (*T. aquaticus*), YP_491259.1 (*E. coli*), EIA15345.1 (*S. aureus*), NP_721232.1 (*S. pneumoniae*), and NP_358573.1 (*S. mutans*), using Clustal Omega with default parameters (Sievers et al. 2011). Asterisks, identical residues; colons, conserved residues; periods, semi-conserved resides; red, residues in *S. pneumoniae* σ^A replaced by bypass mutations.

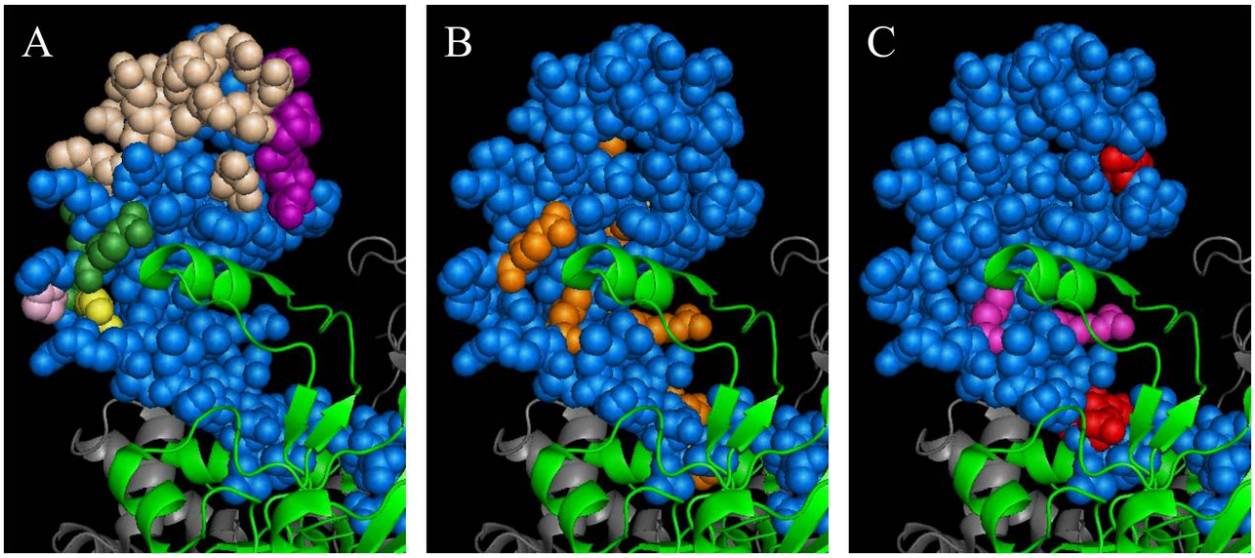


FIGURE 14. Locations of DNA-contacting and protein interaction residues, *comW* bypass residues, and bypass and affinity-affecting residues in region 4 of σ^A in a holoenzyme. Shown is the crystal structure of a holoenzyme from *E. coli* (Bae et al. 2013), Protein Data Bank (PDB) ID 4LHZ. σ^{70}, space-filling blue; ß-subunit, green ribbon; ß', gray ribbon. (A) Residues contacting DNA or regulatory proteins in region 4. Beige, DNA-binding residues; dark green, CAP-interacting residues; pink, FNR-interacting residues; yellow, λCI-interacting residues; purple, PhoB-interacting residues (all according to Campbell et al. (Campbell et al. 2002)). (B) *comW* suppressor residues in region 4. Orange, residues corresponding to *comW* bypass mutations identified in this study. (C) Bypass and affinity residues from *E. coli* and *T. thermophilus* in region 4. Red, residues corresponding to bypass mutations that facilitate the activity of an alternative sigma factor according to Laurie et al. (Laurie et al. 2003); magenta, mutations that reduce σ^A affinity for RNA polymerase core according to Dove et al. (Dove et al. 2003) and Nickels et al. (Nickels et al. 2005).

and thumb are targeted by many transcriptional activator proteins, such as CAP and the λCI repressor, while residues of the little finger interact with PhoB (Fig. 14A) (Campbell et al. 2002). The palm forms a hydrophobic surface that cradles the β flap tip, while the wrist also contacts core residues in the β and β′ subunits. The *comW* bypass substitutions identified here map exclusively to the palm and wrist, surfaces thought to interact with core subunits but not with either promoter DNA or known DNA-binding regulators (Fig. 14B). However, the latter surfaces do contain residues that are important in *E. coli* and *T. aquaticus* for core affinity (Dove et al. 2003; Nickels et al. 2005), as well as residues altered by DksA/ppGpp bypass mutations that facilitate activity of the alternative sigma factor σ^N (Fig. 14C) (Laurie et al. 2003). Colocation of residues altered in *comW* bypass mutants with DksA bypass residues and core affinity residues suggests that ComW may serve in some way to modulate σ^A/σ^X competition in favor of σ^X and that weakening σ^A affinity for core renders this ComW activity less critical for competence development. If σ^A were to have a reduced affinity for core, it is possible that this would manifest in observable changes in growth. We did not observe such changes under standard growth conditions (Fig. 12B), and the literature regarding the residues affecting ppGpp bypass (Laurie et al. 2003) and affinity for core (Dove et al. 2003; Nickels et al. 2005) lacks any observations on corresponding changes in the growth rate. However, there are some mutations in *rpoD* that cause growth defects under certain conditions, such as high temperatures (Grossman et al. 1985) and exposure to ppGpp (Hernandez & Cashel 1995), which we have not yet examined. The alanine residue affected by the remaining two *comW* bypass mutations (*S. pneumoniae* A171 or *E. coli* A415) is buried within the three-dimensional (3D) structure of region 2, where it forms part of the interface between two successive alpha-helices (E399-D417 and K425-Q445). We speculate that disruption of this packing may also weaken the interaction between σ^A and core.

Despite the predominant location of *comW* bypass mutations in areas of σ^A that interact with core subunits, neither the mechanism of action of ComW nor the mechanism bypassing the *comW* requirement is revealed directly by these data. They do not even reveal which of the *comW* phenotypes are relieved in the bypass mutants. The location of affected residues in bypass mutants simply points to σ^A/σ^X competition (Fig. 15), which could be impacted by mechanisms affecting the level of σ^X or by those affecting more directly its interaction with core polymerase. However, it can be thought to "tip the balance" among known mechanisms affecting alternative sigma factor activity if the absence of other suppressor sites is considered. For example, activity of ComW either as an anti-anti-sigma or as an anti-protease-adapter would imply that these partners could be sites for additional effective suppressor mutations. The absence of such sites from our collection of suppressors suggests they are not important targets of ComW activity. In contrast, exclusive targeting of this part of σ^A is easily accommodated if ComW acts to exclude σ^A from polymerase, to promote loading of σ^X onto core polymerase, or to provide extra free core enzyme.

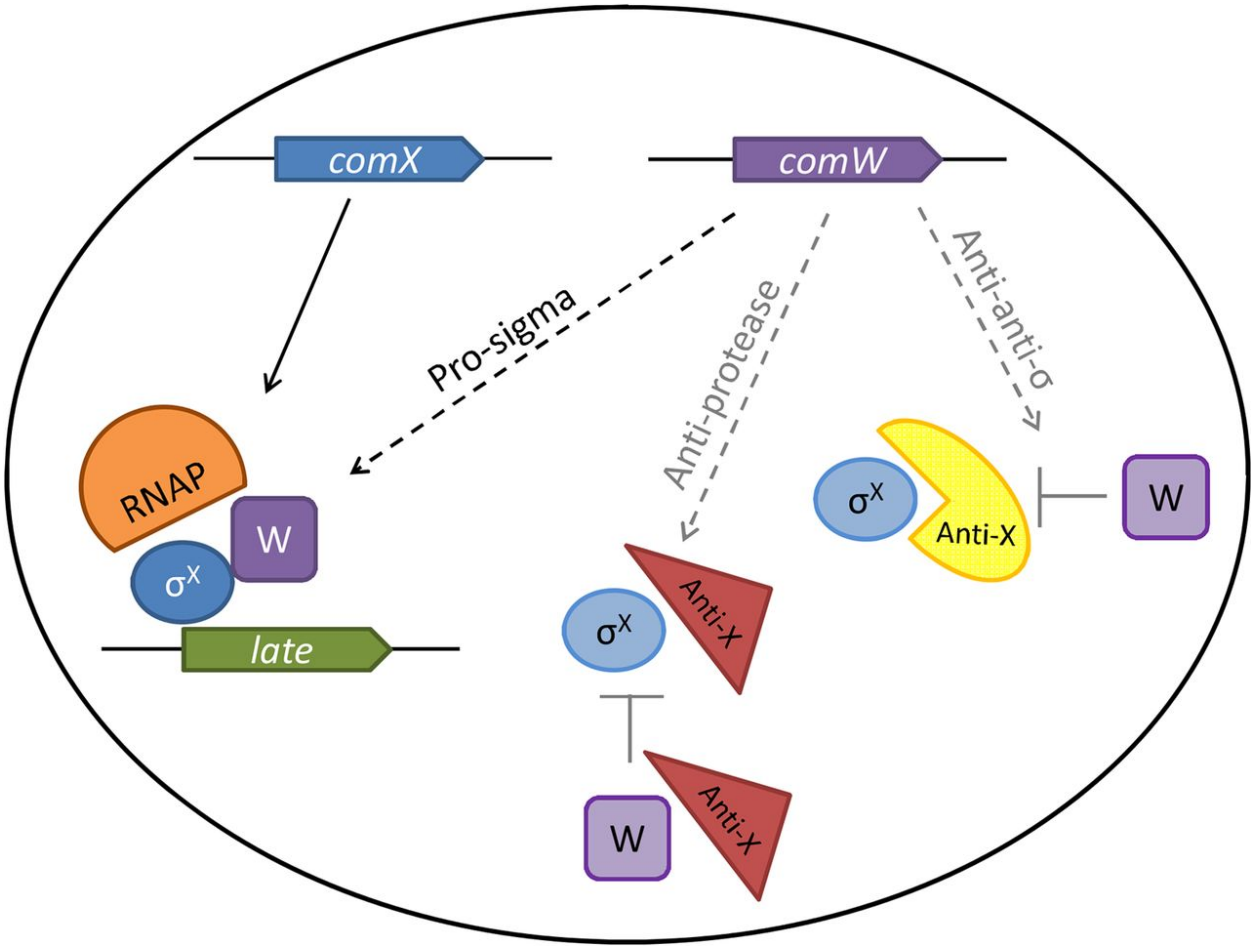


FIGURE 15. Model of potential ComW functions. Possible functions of ComW as a pro-sigma, anti-protease-adapter or as an anti-anti-σ. The pro-sigma function promoting competition with σ^A is in boldface because it is consistent with recovery of bypass mutations exclusively in *rpoD*. Anti-protease and anti-anti-σ functions are lighter because no bypass mutations were in protease subunits/adapters or in other proteins orthologous to anti-σ factors. Dashed lines, possible functions; pentagons, genes; orange half-circle, core RNA polymerase; blue circles, σ^X; purple squares, ComW; yellow, anti-σ^X; red triangles, protease.

IV. Primary sigma factor mutations bypass the ComW requirement for late gene expression during genetic transformation in *Streptococcus pneumoniae.*

4.1 Abstract

Streptococcus pneumoniae is naturally transformable, or able to integrate exogenous DNA into its genome. Competence for genetic transformation in the streptococci depends on an alternative sigma factor, σ^X, for coordinated synthesis of at least 23 proteins, which together establish the competent state. Transient accumulation of high levels of σ^X is coordinated amongst streptococci by transcription factors linked to peptide pheromone signals. σ^X recognizes a non-canonical promoter upstream of late competence genes, such as those required for DNA uptake. In *S. pneumoniae*, elevated σ^X is insufficient for development of full competence without expression of a second competence-specific protein, ComW, which is regulated by the same pheromone circuit that controls σ^X. The Δ*comW* mutant is 10^4-fold deficient in the amount of transformants, 10-fold deficient in the amount of σ^X activity, and 10-fold deficient in the amount of σ^X protein. Previous work identified certain σ^A* mutants that transform well in a Δ*comW* background. To investigate the mechanism of transformation restoration, we measured late gene expression in σ^A* strains. Restoration of late gene expression and competence kinetics was observed in Δ*comW* σ^A* mutants. A lack of late gene shut-off in Δ*comW* mutants points to a potential new role for ComW in competence shut-off. Also, portions of σ^A from *S. mutans* were able to bypass the ComW requirement during transformation. These data are consistent with the idea that ComW increases σ^X access to core RNAP, pointing to a direct role of ComW in σ factor exchange during genetic transformation.

4.2 Introduction

Streptococcus pneumoniae is a Gram-positive opportunistic pathogen found in the human nasopharynx, and it can cause diseases such as pneumonia and meningitis. *S. pneumoniae's* natural competence, or ability to integrate exogenous DNA into its chromosome, provides its major mechanism of rapidly overcoming selective pressure (Chaguza et al. 2015; Johnston, Campo, et al. 2014). The ability to take up and exchange DNA depends on development of the competent state, where quorum sensing (QS) mediated by a small peptide prompts the entire population to become competent in synchrony.

Cells in log-phase are first primed for transformation by a QS mechanism encoded by two genetic loci initially transcribed at a basal level by core RNA polymerase (RNAP) bound to the primary sigma factor (Martin et al. 2013). The loci are *comAB* (Hui & Morrison 1991; Pestova et al. 1996) and *comCDE* (Martin et al. 2013). ComC is a pro-peptide cleaved and exported by ComAB, and the mature peptide is called CSP (competence-stimulating peptide) (Lacks 2004). CSP is sensed by the histidine kinase receptor ComD, which then phosphorylates its cognate response regulator, ComE (Martin et al. 2013; Pestova et al. 1996; Håvarstein et al. 1996; Ween et al. 1999). Phosphorylated ComE (ComE~P) binds at an inverted repeat upstream of the promoters and upregulates the activity of *comAB, comCDE,* and six more operons that are transcribed only at competence (Martin et al. 2013; Martin et al. 2010; Oggioni & Morrison 2008), called the early genes. Induction of the early genes incites a positive feedback loop that increases the level of CSP produced.

Another early gene is *comX*, encoding the only known streptococcal alternative sigma factor, σ^X (Lee & Morrison 1999). σ^X forms a holoenzyme with RNAP and recognizes non-canonical promoter sequences, or cin-boxes (competence-induced), upstream of late competence

genes (Luo et al. 2003; Luo & Morrison 2003; Peterson et al. 2004; Campbell et al. 1998). The ComABCDE/CSP system of competence induction is maintained in the mitis and anginosus groups of streptococci, whereas other streptococcal groups utilize a different set of QS genes, *comR* and *comS* (Håvarstein 2010). Strict regulation of σ^X is imperative in the streptococci as some competence genes are responsible for fratricidal proteins and associated immunity proteins. In addition to transcriptional regulation by QS, σ^X stability and activity relies on a small protein, ComW (Luo et al. 2004; Sung & Morrison 2005). *comW* sequences have only been found in eleven species of streptococci, but every streptococcus species has σ^X (Johnston, Martin, et al. 2014), suggesting that species without ComW have other mechanisms of regulation of σ^X that do not require ComW.

comW was first identified as a gene whose expression was CSP-inducible and required for transformation (Bartilson et al. 2001). ComW was subsequently implicated in the full σ^X response during transformation (Luo et al. 2004), ectopic coexpression of σ^X and ComW was enough to restore the amount of transformants to 80% of WT levels. ComW was also necessary for σ^X activity (Sung & Morrison 2005), and for σ^X protection from degradation by the ATP-dependent ClpP protease. Δ*comW* Δ*clpP* strains had restored amounts of σ^X, but were unable to transform despite having the native amounts of σ^X (Sung & Morrison 2005). A recent study showed that several single base substitutions in σ^A, the primary σ factor, can bypass the ComW requirement during transformation, and that these substitutions map to residues implicated in σ^A affinity for core RNAP (Tovpeko & Morrison 2014). This data suggested that ComW acts at the point of σ switching during transformation, from σ^A to σ^X.

While several Δ*comW* mutant phenotypes are known, its overall function remains unclear. First, Δ*comW* mutants are 10^4-fold deficient in transformants (Luo et al. 2004). Second, Δ*comW*

mutants have 10% of WT late gene transcription (Sung & Morrison 2005). Third, Δ*comW* mutants display 10% of WT amounts of σ^X protein (Sung & Morrison 2005). Fourth, Sung and Morrison genetically separated the activity and stability functions of ComW, the C-terminus was responsible for σ^X stability, while the N-terminus of ComW was responsible for both stability and activity (Sung & Morrison 2005). Lastly, single base substitutions in σ^A, called σ^A*, bypass the ComW requirement for transformation (Tovpeko & Morrison 2014), but the mechanism of restored transformation is not known. To attempt to elucidate the mechanism of restored transformation, we first asked whether late gene expression is restored in Δ*comW* σ^A* mutants.

In the present study, we show that late gene expression restoration correlates with transformation restoration in Δ*comW* σ^A* strains. We also report restored competence kinetics in these strains. We observed the persistence of late gene expression in all Δ*comW* mutants, suggesting a new function for ComW in late gene shut-off. Lastly, we show that a portion of the *S. mutans* WT σ^A is a suppressor of ComW in *S. pneumoniae*. Taken together, these data support ComW's role in increasing σ^X access to core RNA polymerase.

4.3 Results

4.3.1 Late gene expression is restored in Δ*comW* σ^A* suppressor strains.

Transformation requires the cooperation of many proteins. In the *comW* mutant, the amount of transformants is 0.01% of the WT. In Δ*comW* σ^A* strains, the yield of transformants was restored up to 10% of the WT level but the exact mechanism of this restoration is unclear. Because transformation depends on σ^X activity to initiate transcription of late genes, it is conceivable that σ^A* restores late gene expression and allows transformation to occur. To test this hypothesis, we asked whether late gene expression was restored in Δ*comW* σ^A* mutants. Each σ^A* strain had a different mutation in σ^A, the primary σ factor, causing one of six amino acid changes (A171T,

A171V, R316C, R316H, R355H, L363F), representing the range of the amino acid mutations found in the suppressor screen that restore the amount of transformants in *comW* mutants (Tovpeko & Morrison 2014).

To monitor late gene expression in Δ*comW* σ^{A}*strains, we measured the amount of LacZ produced by a *lacZ* fusion to the promoter of *ssbB*, a late gene responsible for protecting the newly imported DNA (Attaiech et al. 2011). *ssbB* is transcribed from a cin-box recognized by σ^{X} bound to core RNAP (Peterson et al. 2004; Campbell et al. 1998). Since the cin-boxes are similar amongst the late genes, *ssbB* expression is taken as representative of late gene expression levels. *ssbB* expression in six σ^{A}*, *comW* strains was measured over a period of 80 minutes after induction of competence with CSP (Fig. 16).

In the WT, *ssbB* expression was not detected at the time of addition of CSP (Fig. 16). However, at 20 – 30 minutes after CSP induction there was peak expression, with a decrease of expression beginning at 30 minutes. The decrease of expression reflects the shut-off of late gene expression, proteolytic decay of σ^{X}, and the end of competence. The *comW* mutant had a much lower level of late gene expression than the WT, which continued beyond 80 minutes at a reduced rate. At 20 minutes post-CSP induction in every Δ*comW* σ^{A}* strain, there was a significant increase of late gene expression relative to the Δ*comW* control, indicating restoration of σ^{X} activity. In strains with $\sigma^{A*R316H}$, $\sigma^{A*L363F}$, and $\sigma^{A*R355H}$, expression was roughly 20% that of the WT. σ^{A}* strains with $\sigma^{A*A171V}$, $\sigma^{A*A171T}$, and $\sigma^{A*R316C}$ had 70-80% restored *ssbB* expression at 20 minutes. The pattern of increase in *ssbB* and restoration to 20% to 80% of WT levels in Δ*comW* σ^{A}* mutant strains indicates that the suppressor at least acts to restore σ^{X} activity.

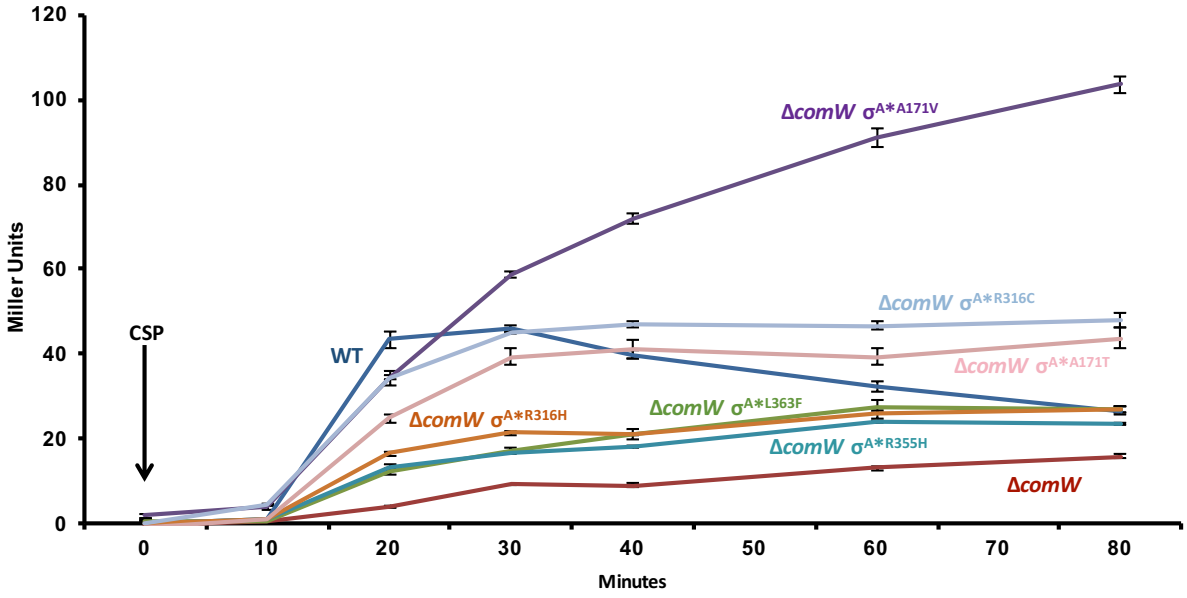


Figure 16: **Late gene expression in Δ*comW* σ^{A}* strains after addition of CSP**. Late gene expression patterns were measured in WT (σ^{AWT}), *comW* (σ^{AWT}), and six Δ*comW* σ^{A}* mutant strains ($\sigma^{A*A171T}$, $\sigma^{A*A171V}$, $\sigma^{A*R316C}$, $\sigma^{A*R316H}$, $\sigma^{A*R355H}$, $\sigma^{A*L363F}$) after CSP induction using a *lacZ* reporter at gene *ssbB*. P_{ssbB} driven transcription of *lacZ* was measured by β-galactosidase activity (Miller units) in samples harvested every 10 minutes after CSP induction. Each strain is indicated by its σ^{A} mutation. WT and *comW* have the WT copy of σ^{A}. Error bars measure standard deviation of three biological replicates. Bold black arrow, addition of CSP.

4.3.2 Late gene expression is persistent in Δ*comW* σ^{A}* strains.

In the WT, there is a decrease in the rate of late gene expression, previously observed (Piotrowski et al. 2009; Luo et al. 2004) and is expected due to shut-off of competence and the end of late gene expression (Martin et al. 2013; Mirouze et al. 2013). This rate decrease was not observed in the Δ*comW* mutant, *ssbB* expression persisted at 40 minutes after CSP induction, and this same pattern was observed in all Δ*comW* σ^{A}* strains (Fig. 16). The rate of *ssbB* synthesis can be measured by slope of expression between two time points. A negative or positive rate can indicate increases or decreases in expression, respectively. In Δ*comW* σ^{A}* strains after 40 minutes, *ssbB* expression continued, and was 15-30% of the original rate of expression, as measured by the slope of the expression values at those times (Table IX). It is possible that this persistence is due to the lack of *comW*, and one of the functions of ComW could be to deliver σ^{X} to protease to ensure proper shut-off of competence. In this case, we hypothesize that we would not expect to see persistence, but normal competence shut-off in σ^{A}* strains that are $ComW^{+}$. If the decrease of *ssbB* expression at 40 minutes after induction relies on ComW, we would expect to see normal induction and decrease of *ssbB* expression similar to WT in σ^{A}* strains that are $ComW^{+}$.

To monitor late gene expression in σ^{A}* strains, expression was measured for 80 minutes, and compared to their corresponding strain Δ*comW* σ^{A}* strain (Fig. 17). We observed that $ComW^{+}$ strains had negative rates of expression, while Δ*comW* σ^{A}* strains had positive rates, indicating persistence of expression (Fig. 17, Table X). The data implies that ComW is responsible for the proper regulation and shut-off of σ^{X} activity, because late gene expression is normal regardless of the σ^{A}* mutation if ComW is present in this strain, and strains lacking *comW* all have persistent *ssbB* expression.

Table IX. Late gene expression rates at time of induction (10-20 min) and expected shut-off (40-60 min).

	ComW[b]			
	+		-	
σ^A allele[a]	Induction[c]	Shut-off[d]	Induction[c]	Shut-off[d]
WT	4.2	-0.4	0.4	0.2
A171V	2.9	-0.4	3.0	1.0
R316H	5.0	-0.5	1.6	0.2
R355H	4.1	-0.5	1.3	0.3
L363F	5.2	-0.5	1.1	0.3

[a]Allele of σ^A in strain tested

[b]Presence of absence of ComW in strain tested

[c]Rate of P_{ssbB} expression as Miller units/mL, measured by slope of expression from 10 to 20 minutes

[d]Rate of P_{ssbB} expression as Miller units/mL, measured by slope of expression from 40 to 60 minutes

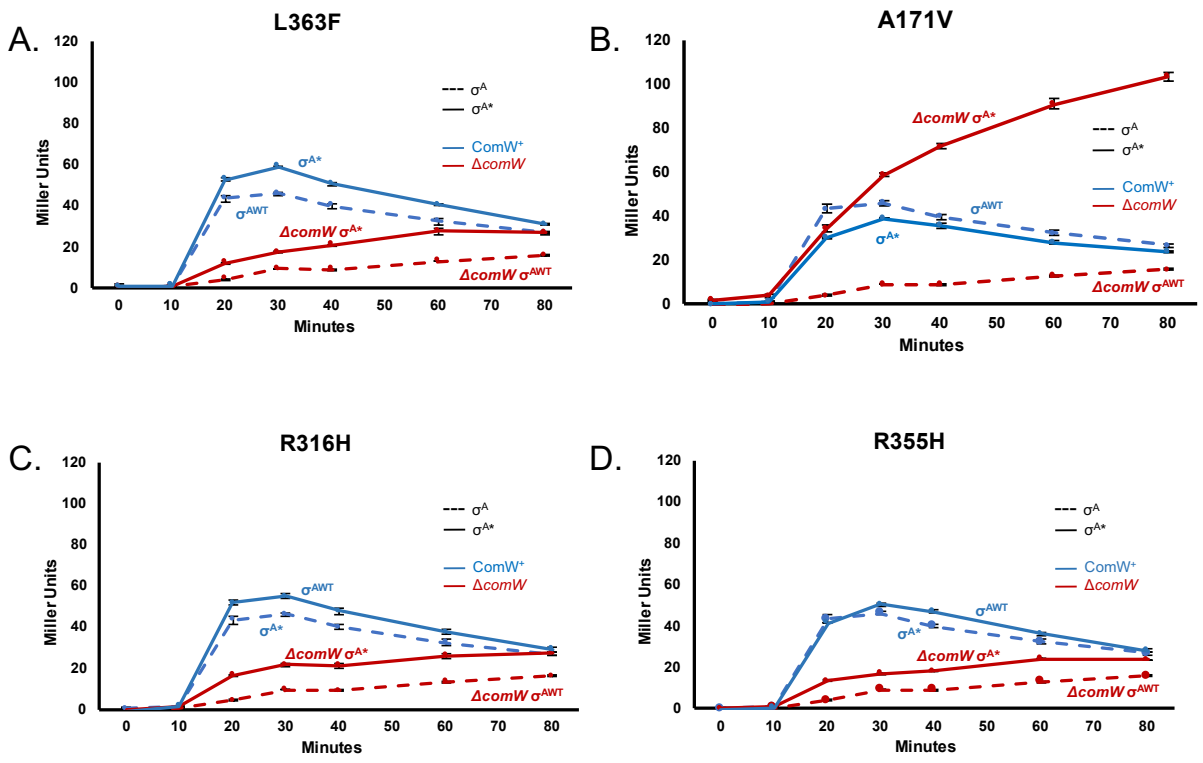


Figure 17: Late gene expression in σ^{A}* strains with and without ComW after CSP addition. Late gene expression patterns after CSP induction using a *lacZ* reporter at gene *ssbB*. β-galactosidase activity (Miller units) was measured in samples harvested every 10 minutes after CSP induction were measured in WT (σ^{AWT}), Δ*comW* ($\sigma^{AWT)}$, and σ^{A}* strains in the presence or absence of ComW. Each panel examines a different single amino acid substitution in σ^{A}, denoted as the WT amino acid, the position of the amino acid change, and the amino acid identified in the suppressor at that position, (A) $\sigma^{A}*^{L363F}$; (B) $\sigma^{A}*^{A171V}$; (C) $\sigma^{A}*^{R355H}$; (D) $\sigma^{A}*^{R316H}$. ---, strains with WT σ^{A}; ——, strains with σ^{A}* suppressor mutations; red, strains with *comW* deletions; blue, strains with WT *comW*. Error bars measure standard deviation of experiments performed in triplicate.

4.3.3 Competence kinetics are normal in σ^{A}* mutants.

One of the key components of transformation in *S. pneumoniae* is the rapid shut-off of competence by negative feedback regulators such as DprA (Martin et al. 2013; Mirouze et al. 2013) and σ^{X} degradation by the ATP-dependent ClpP protease (Robertson et al. 2002; Chastanet et al. 2001; Luo & Morrison 2003). However, we have observed that ComW may also be responsible for proper shut-off of σ^{X} activity (Fig. 17, Table IX), potentially allowing for extended transformation in strains that have extended late gene expression. In the Δ*comW* σ^{A}* strains, we expect to also see transformants past the time of expected shut-off, because of the observed persistent late gene expression.

To monitor transformants over time in Δ*comW* σ^{A}* mutant strains, we induced competence and evaluated the amount of transformants produced by 5-minute DNA exposure after CSP induction (Fig. 18). Just as with *ssbB* expression, transformants peaked at 20 minutes after CSP induction. After 30 minutes, the level of transformants dropped off sharply and were undetectable for all strains; except Δ*comW* $\sigma^{A*A171V}$. This strain still transformed past 60 minutes, but with 5-fold lower transformants than during the time of peak late gene expression. This was consistent with the earlier finding that late gene expression does not seem to shut off in this strain, and has a higher rate of *ssbB* expression than any other Δ*comW* σ^{A}* strain (Fig. 16, 17B, Table IX). It is interesting that transformation persists in the presence of ClpP, the protease responsible for degradation of σ^{X}. Despite the presence of ClpP, transformation persists if $\sigma^{A*A171V}$ is present.

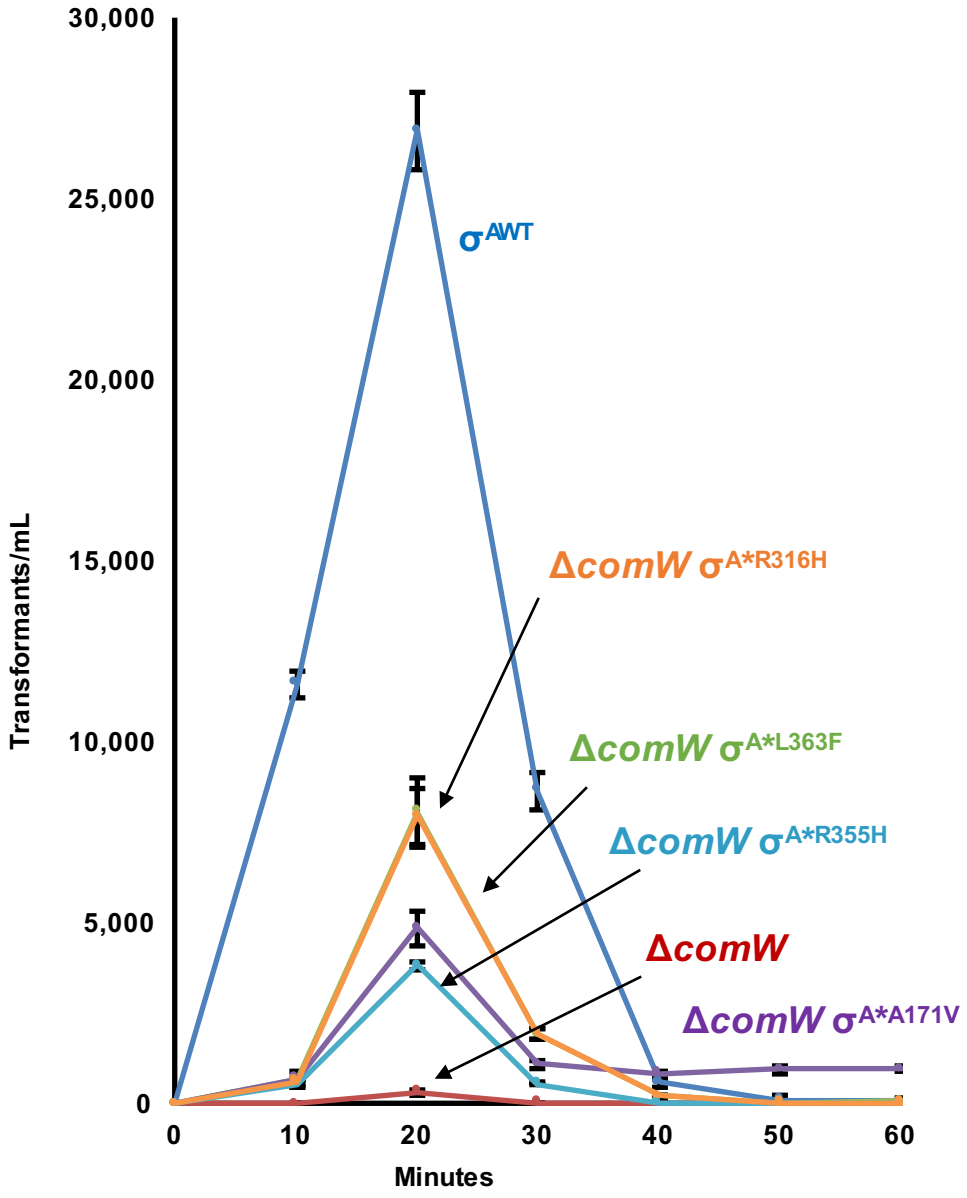


Figure 18. Competence kinetics in Δ*comW* σ^{A}* strains after addition of CSP. The number of transformants/mL after 5-minute exposure to DNA was measured in WT (σ^{AWT}), Δ*comW* (σ^{AWT}), and six Δ*comW* σ^{A}* mutant strains ($\sigma^{A*A171T}$, $\sigma^{A*A171V}$, $\sigma^{A*R316C}$, $\sigma^{A*R316H}$, $\sigma^{A*R355H}$, $\sigma^{A*L363F}$) after CSP induction and 5-minute exposure to 5 ng/mL DNA at each time point. Each strain is indicated by its σ^{A} mutation. WT and *comW* have the WT copy of σ^{A}. Error bars measure standard deviation of experiments performed in triplicate. Bold black arrow, addition of CSP.

4.3.4 The WT σ^A from *Streptococcus mutans* is a suppressor of the ComW requirement in *S. pneumoniae.*

comW sequences have only been identified in eleven species in the mitis and anginosus groups, even though every single streptococcus species contains a σ^X (Johnston, Martin, et al. 2014), and most species have homologs of genes required for competence (Peterson et al. 2004; Reck et al. 2015; Vickerman et al. 2007; Rodriguez et al. 2011; Woodbury et al. 2006; Fontaine et al. 2010). It is interesting that σ^X is ubiquitous but ComW seems to be restricted to the mitis and anginosus groups, and that some streptococci such as *S. mutans* are highly competent without ComW. We hypothesized that σ^A from ComW-less streptococcal species could be a suppressor of ComW in *S. pneumoniae.*

We compared the amino acid sequences of σ^A from transformable streptococcal species lacking ComW; *S. thermophilus, S. equinus,* and *S. mutans* to *S. pneumoniae*. We focused on regions 2-4, the locations of ComW suppressor residues in *S. pneumoniae*, to identify any residues that may be possible suppressors of the Δ*comW* mutant phenotype (Fig. 19A, yellow background). Several residues were not conserved amongst the species, we focused on *S. mutans* because it had the highest number residues differing from *S. pneumoniae*. In total, 6 residues were not conserved between *S. mutans* and *S. pneumoniae*, with two differing residues immediately adjacent to residues implicated in ComW bypass in *S. pneumoniae* (Tovpeko & Morrison 2014). We predicted that the six residues that are different in *S. mutans* will be partial suppressors of the *comW* defect during transformation (Fig. 19A, blue background).

To test whether the *S. mutans* σ^A could restore the transformation phenotype in the *S. pneumoniae comW* background, we created a chimeric σ^A that retained region 1 from *S. pneumoniae* but had regions 2-4 from *S. mutans,* a 6 residue difference in total (Fig 20B). We

A.

```
Chimera
S.pneumo     KGFKFSTYATWWIRQAITRAIADQARTIRIPVHMVETINKLVREQRNLLQELGQDPTPEQ    238
S.mutans     KGFKFSTYATWWIRQAITRAIADQARTIRIPVHMVETINKLVREQRNLLQELGQDPTPEQ    240
S.thermo     KGFKFSTYATWWIRQAITRAIADQARTIRIPVHMVETINKLVREQRNLLQELGQDPTPEQ    238
S.equin      KGFKFSTYATWWIRQAITRAIADQARTIRIPVHMVETINKLVREQRNLLQELGQDPTPEQ    238
             ************************************************************

Chimera                                                          V
S.pneumo     IAERMDMTPDKVREILKIAQEPVSLETPIGEEDDSHLGDFIEDEVIENPVDYTTRIVLRE    298
S.mutans     IAERMDMTPDKVREILKIAQEPVSLETPIGEEDDSHLGDFIEDEVIENPVDYTTRVVLRE    300
S.thermo     IAERMDMTPDKVREILKIAQEPVSLETPIGEEDDSHLGDFIEDEVIENPVDYTTRVVLRE    298
S.equin      IAERMDMTPDKVREILKIAQEPVSLETPIGEEDDSHLGDFIEDEVIENPVDYTTRVVLRE    298
             ******************************************************:*****

Chimera          V                                   D                   H
S.pneumo     QLDEILDTLTDREENVLRLRFGLDDGKMRTLEDVGKVFNVTRERIRQIEAKALRKLRQPS    358
S.mutans     QLDEVLDTLTDREENVLRLRFGLDDGKMRTLEDVGKVFDVTRERIRQIEAKALRKLRHPS    360
S.thermo     QLDEVLDTLTDREENVLRLRFGLDDGKMRTLEDVGKVFNVTRERIRQIEAKALRKLRHPS    358
S.equin      QLDEVLDTLTDREENVLRLRFGLDDGKMRTLEDVGKVFNVTRERIRQIEAKALRKLRHPS    358
             ****:**********************************:******************:**

Chimera         Q    V
S.pneumo     RSKPLRDFIED   369
S.mutans     RSKQLRDFVED   371
S.thermo     RSKQLKDFIED   369
S.equin      RSKQLRDFIED   369
             *** *:**:**
```

B.

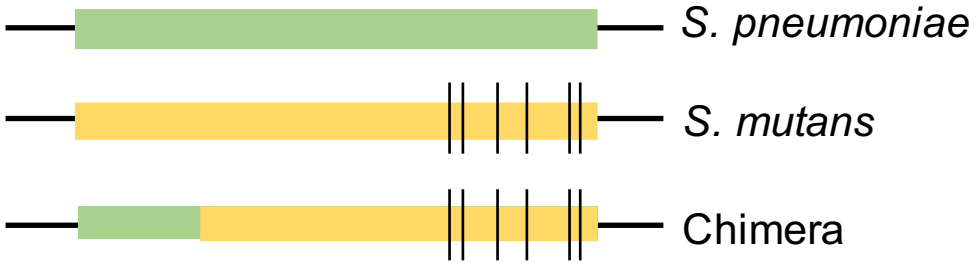


Figure 19: Alignment of σ^A homologs from *S. pneumoniae* and *S. mutans* and construction of chimeric σ^A. (A) *S. pneumoniae* σ^A residues 117 to 369 (accession number NP_721231.1) were aligned with corresponding residues from *S. mutans* (NP_358573.1), *S. thermophilus* (AKH33903.1), and *S equinus* (EFM27146.1), using Clustal Omega with default parameters (Sievers et al. 2011). Asterisks, identical residues; colons, conserved residues; yellow background, residues with at least one species having a different amino acid from *S. pneumoniae*; green background, residues altered in individual ComW bypass mutants; blue background, residues different from *S. pneumoniae* in chimeric σ^A. Purple bar, region 2 residues; Pink bar, region 3 residues; orange bar, region 4 residues. (B) *S. mutans* σ^A chimera. After colony selection and strain verification, the resulting chimera had region 1 from *S. pneumoniae* and regions 2-4 from *S. mutans*. Yellow rectangles, σ^A regions from *S. mutans*; green, σ^A regions from *S. pneumoniae*; black vertical lines, divergent residues; black horizontal lines, genome.

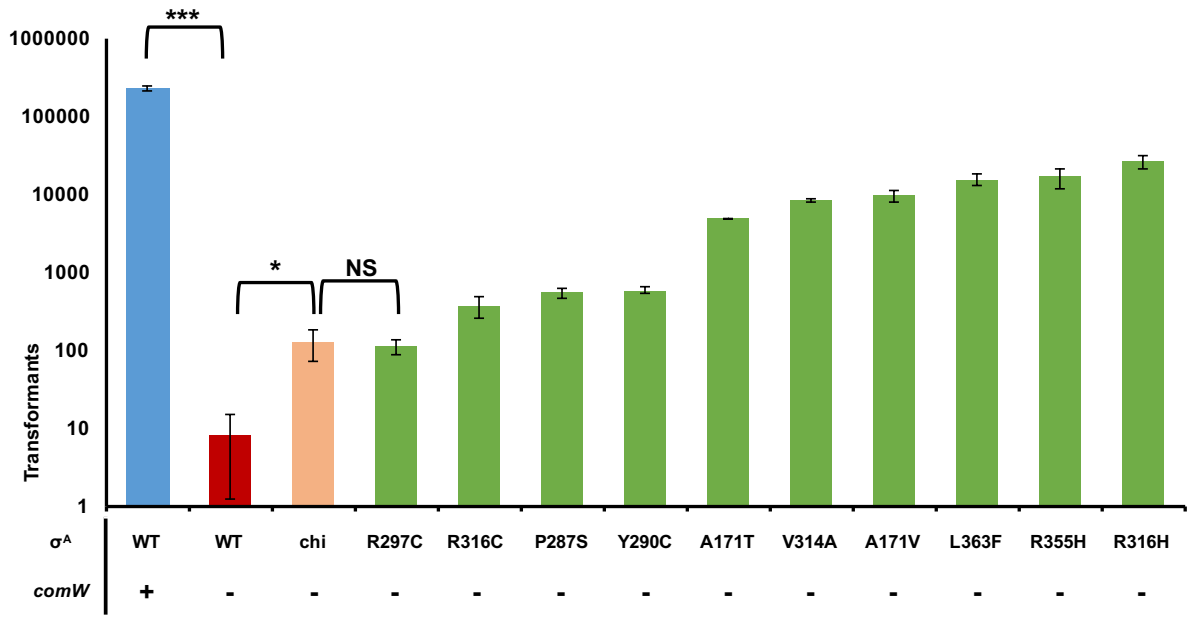


Figure 20. Suppression of *comW* phenotype by portions of *S. mutans* σ^A. Transformation efficiency was measured in WT (σ^{AWT}), *comW* (σ^{AWT}), a Δ*comW* chimeric σ^A (σ^{A*chi}) strain with six amino acids from *S. mutans*, and ten σ^{A*} suppressor strains, denoted by the mutation in σ^A (Tovpeko & Morrison 2014). Statistical significance was measured using Student's two-tailed T-test. NS, not significant; *, $p < 0.05$; ***, $p < 0.001$.

transformed the chimeric σ^{A} construct (σ^{Achi}) into *S. pneumoniae*, and removed *comW* from the strain, followed by the standard transformation assay. The σ^{Achi} restores transformation efficiency significantly above the level of the Δ*comW* mutant in *S. pneumoniae* (Fig. 5). We also compared the σ^{Achi} level of transformation restoration to that of σ^{A}* bypass suppressors. The level of transformation restoration by one of the ComW bypass strains, ALT5 (σ^{AR297C}), is not significantly different from that of σ^{Achi}. This shows that the *S. mutans* σ^{A} is a partial suppressor of ComW in *S. pneumoniae*. These data suggest that the need for ComW in *S. pneumoniae* may be overcome in other streptococci by their respective σ^{A}, and that some streptococci have mechanisms of σ^{X} regulation that do not involve ComW.

4.4 Discussion

To determine the mechanism of transformation restoration in ComW suppressors, we investigated late gene expression in Δ*comW* σ^{A}* strains and found a correlation between the restoration of competence and the restoration of late gene expression. We also show that the persistence of late gene expression is due to the lack of ComW. This is remarkable because it reveals that ComW may play a role in both the induction and shut-off of late gene expression. Furthermore, we determined that portions of the WT in σ^{A} from *S. mutans* bypasses the ComW requirement for transformation in *S. pneumoniae*. Measuring the patterns of late gene expression in ComW bypass strains hinted at a new potential role for ComW during natural transformation in *S. pneumoniae*.

ComW was reported to have many roles in transformation, such as the promotion of competence, and the stability and activity of σ^{X}, so we wanted to pinpoint the mechanism of transformation restoration in Δ*comW* σ^{A}* strains. Despite almost complete restoration of late gene expression in strains in some Δ*comW* σ^{A}* strains at 25 minutes, transformation was 10-fold below

the WT (Fig. 20) (Tovpeko & Morrison 2014). Further study is needed to understand the correlation between the small differences in late expression but large fold differences in transformation efficiency. Many studies analyzing mutations in primary σ factors have found increased transcription from alternative σ factor promoters and evidence of σ factor competition (Laurie et al. 2003; Yin et al. 2013; Lee & Helmann 2014), suggesting that ComW acts to modulate competition between σ^A and σ^X. The increase in late gene expression does not rule out that the amount of σ^X is also restored in Δ*comW* σ^A* mutant strains. It is possible that transformation is also restored because of an increase in the amount of σ^X, which in turn results in an increase in late gene expression. This would need to be tested for confirmation by measuring σ^X amounts in all σ^A* strains. Studies exploring σ factor competition have also found that mutant phenotypes can be restored by overproduction of a σ factor (Tu et al. 2006; Farewell et al. 1998). We predict that the amount of σ^X in σ^A* strains will be restored above the level of the *comW* mutant.

The implication of ComW's role in the shut-off of late gene expression is futher complicated by the technique used to measure expression rates. LacZ has a turnover rate that is likely different than that of the mRNA it is measuring. To confirm that ComW is responsible for the shut-off of late gene expression, other techniques such as qRT-PCR would need to be employed to measure message during late competence.

It is interesting that the rate of late gene expression restoration varies from 10% to 80% of WT levels, but the transformation efficiency strains remains at 10% of the WT level. This could indicate that there is some critical threshold that is needs to be achieved for the level of transformation to match WT, and that threshold is higher than the 80% shown in two strains ($\sigma^{A*A171V}$ and $\sigma^{A*R316C}$). The idea of a critical threshold for transformability is supported by our observation that the frequency of transformants was restored in σ^A* strains, indicating that the

same portion of the population as WT is transforming. However, since the actual amount of transformants is only 10% in σ^{A}* strains, this indicates that transformation was imperfectly restored to 90% of the transforming population, perhaps because of a threshold required for a critical amount of late gene expression.

The inference of a threshold for late gene expression is further complicated by the fact that only one late gene was investigated in this study. Even though all the cin-boxes upstream of the late genes are highly conserved (Peterson et al. 2004), gene expression and protein levels of additional late genes needs to be investigated to confirm the presence of a threshold of expression required for transformation. Furthermore, it appears that ComW is also responsible for the proper shut-off of late gene expression, as evidenced by the persistence of some late gene expression in strains lacking *comW*. This could add another role to the list of functions previously identified for ComW in the function of σ^{X} during transformation. There have been reports of genes transcribed by alternative σ factors that negatively regulate that σ (Straus et al. 1990), and it is possible that a late gene is responsible for ComW's role in late gene shut-off.

Despite the reliance on ComW for the normal σ^{X} response during transformation in *S. pneumoniae*, many species of streptococci are competent without a protein resembling ComW being identifiable in their genomes (Reck et al. 2015; Fontaine et al. 2010; Morrison et al. 2013). When we mapped the ComW bypass residues from the previous study and the six residues that differ in *S. mutans* (Fig. 21), three of the six face the same critical flap of the β subunit that is responsible for σ^{A} region 4 affinity for RNAP (Dove et al. 2003; Campbell et al. 2002). This suggests that the *S. mutans* residues we tested may also be important for σ^{A} affinity for RNAP, and could explain the bypass effect observed. The three resides that do not face this flap may have no effect on ComW bypass, and any of the three that do face the flap may be solely conferring the

bypass effect. The residues would need to be tested individually to confirm whether any one is responsible for ComW bypass. It is possible that other transformable streptococcal species overcome the σ^X requirement for ComW by residues in σ^A that vary from those present in *S. pneumoniae*. However, it is also possible that the σ^X in other species has a higher affinity for RNAP than that of *S. pneumoniae*, and this would need to be confirmed by testing various forms of σ^X from transformable streptococci. Overall, our data suggest that ComW's role is to act as a σ factor activator, increasing σ^X access to RNAP and an additional function may be to ensure proper shut-off of σ^X activity (Fig. 22). One example of σ factor activators exists, called Crl, but only found in Gram-negative organisms (Banta et al. 2013). ComW's role as an accessory protein in σ^X activity could be a novel mechanism of alternative σ function in Gram-positive bacteria.

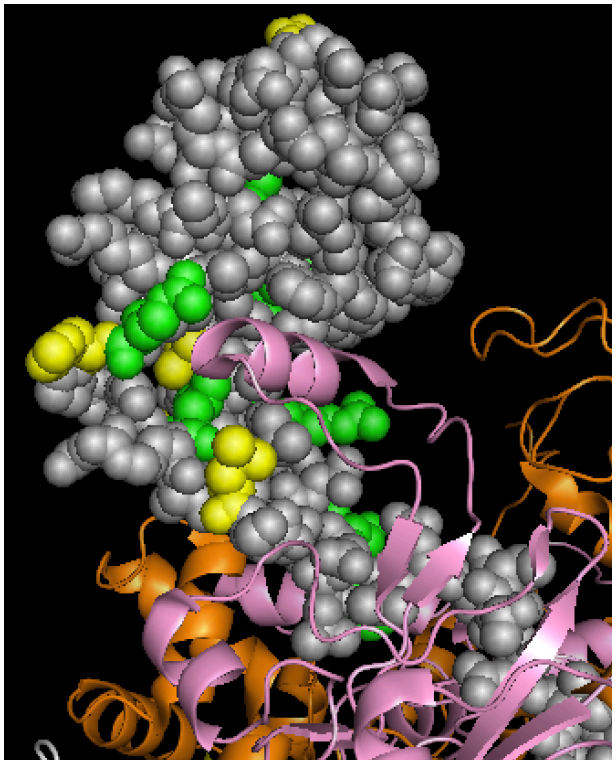


Figure 21. Locations of ComW bypass suppressor substitutions and divergent *S. mutans* residues that bypass ComW in region 4 of σ^A in a holoenzyme. Shown is the crystal structure of a holoenzyme from *E. coli* (Bae et al. 2013), Protein Data Bank (PDB) ID 4LJZ. Residues identified as ComW bypass residues are in green, *S. mutans* residues different from *S. pneumoniae* shown in yellow. Space filling gray, σ^A; pink ribbon, β-subunit; orange ribbon, β'.

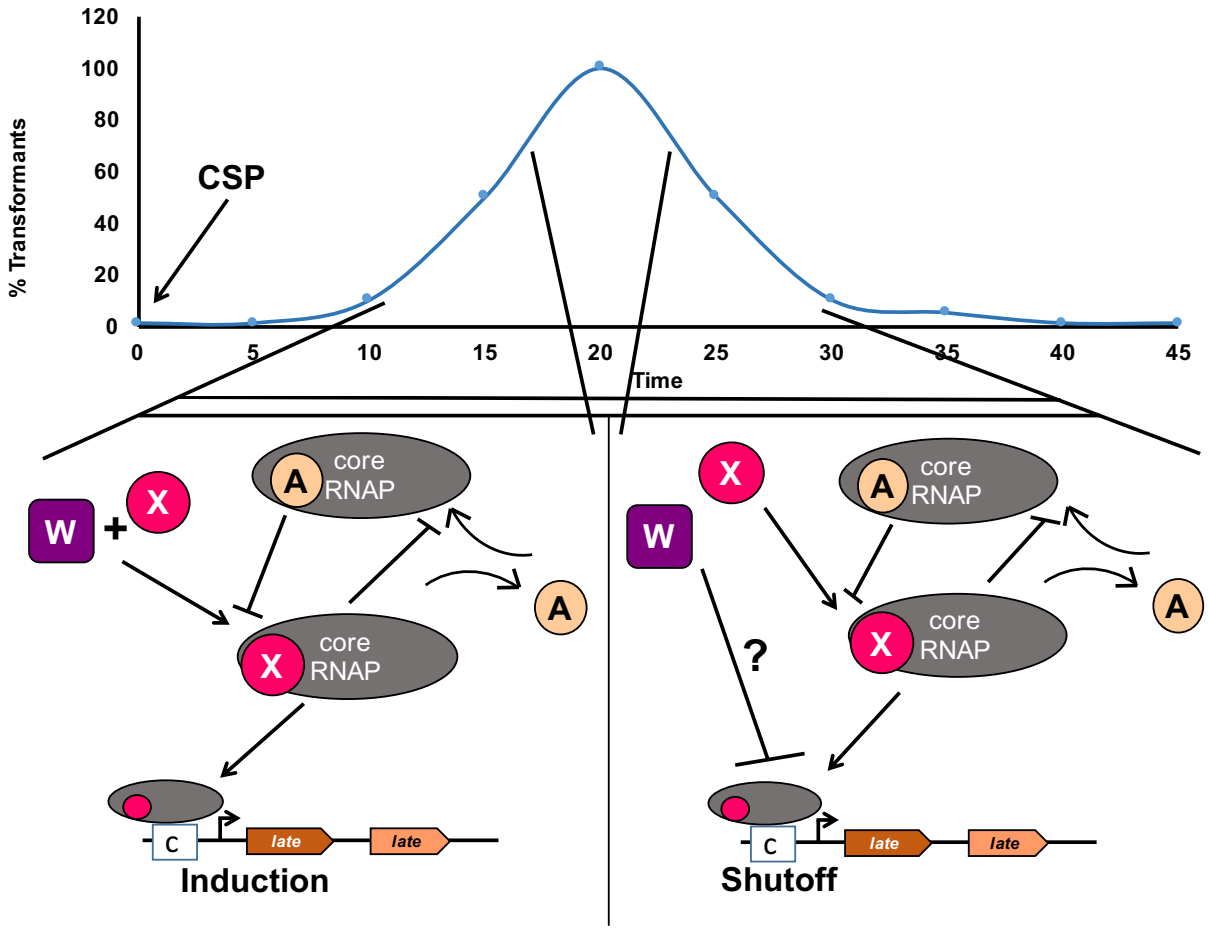


Figure 22. Model of the potential action of ComW during transformation in *S. pneumoniae*. From CSP induction until the time of peak σ^X activity (20 minutes), ComW functions to promote σ^X access to core RNA polymerase to allow for transcription of the late genes necessary to complete transformation. After peak σ^X activity, ComW may function to assist in the shut-off of late gene expression. The question mark points to the unknown effect of ComW on the shut-off of late gene expression. The lack of shut-off of late gene expression in Δ*comW* mutants points to ComW having a function in controlling both the upregulation and downregulation of late genes during transformation. CSP, competence stimulating peptide addition; purple squares, ComW; pink circle, σ^X; beige circle, σ^A; gray oval, core RNAP; boxed c, cin-box; brown pentagons, late genes; bent arrow, promoter.

V. Strategies for genome modification in *Streptococcus pneumoniae* by natural genetic transformation: markerless genome editing and dependence of transformation efficiency on donor.

5.1 Abstract

Transformation efficiency in some strains of *Streptococcus pneumoniae* can reach close to 100%, presenting an attractive opportunity for high efficiency genome editing that does not require markers, selection, or result in unwanted changes to the genome. Parameters that affect transformation efficiency in *S. pneumoniae* include temperature, type of donor (genomic DNA, PCR amplicons, plasmids), homology of donor to recipient, and amount of donor DNA. Several genetic tools such as the Janus cassette and the self-deleting Cheshire cassette have been created to improve genome editing in *S. pneumoniae.* I investigated the dependence of transformation efficiency on the amount and structure of donor DNA for both genomic DNA donor and PCR amplicon donors. I also monitored the dependence of transformation efficiency on the amount of homologous regions in PCR amplicons, ranging from one-base substitutions to 1.2-kb cassettes. I observed that transformation efficiency increased with the amount of donor DNA for both PCR and genomic donors, but the saturating amount varied by ~100-fold. I demonstrate that the near 100% transformation efficiency of PCR amplicons can be used to perform markerless, scar-less, and selection-less editing on an essential gene.

5.2 Introduction

One of the best studied characteristics of *S. pneumoniae* is the natural ability to transform, taking up exogenous DNA and integrating it into its genome. This natural capability of the organism to modify its genome, called competence, allows for the organism's genomic plasticity. Transformation has long been studied in *S. pneumoniae,* first observed in this same bacterium in 1928 (Griffith 1928). Transient competence development is achieved through the export and sensing of a peptide pheromone (reviewed in (Johnston, Campo, et al. 2014)). Use of synthetic peptide allows for the control of the timing of transformation without the need to wait for endogenous competence development. The ability of the bacterium to integrate DNA into its genome based on homology has allowed for extensive genetic study, and many parameters of transformation efficiency based on donor DNA have been measured.

Transformation efficiency in *S. pneumoniae* depends on many different factors including temperature, the size of the donor molecule, and the topology of the donor molecule. Transformation efficiency is dependent on temperature and the transformation frequency for multiple transformants was lower than the transformation efficiency for any individual marker (Ghei & Lacks 1967). There is a linear dependence of transformation on the size of donors from 1- to 10-kb (Cato & Guild 1968). Similar transformation parameters were found for another naturally transformable bacterium, *B. subtilis* (Morrison & Guild 1972; Contente & Dubnau 1979). More studies revealed that the rate of transformation decreases when the amount of mismatch between donor and recipient increases (Claverys & Lacks 1986). Transformation efficiency decreases with the decrease in size of donor, with markers that only have one homologous arm, and with reductions in homology of the donor to the recipient (reviewed in (Lacks 1978; Mortier-

Barriere et al. 1997)). The above mentioned studies used genomic donors, or preparations of donor DNA from the entire genome of a strain containing the desired marker.

More recent studies began exploring transformation dependence on synthetic donors, either plasmids or linear PCR amplicons. Transformation efficiency does not depend on the genome site targeted but does depend on the size of the insert on a plasmid, transformation efficiency decreases with the decrease of insert size (Lee et al. 1998). For PCR amplicons consisting of cassettes between flanking homologous arms, increases in flanking arm length increased transformation efficiency, but varying either homologous arm decreased transformation efficiency (Lau et al. 2002). Recent studies increased the ability to edit the genomes of streptococci while reducing the introduction of unrelated sequences and potential polar effects by using large donor amplicons and co-transformation of a selectable marker with a second unselected marker (Dalia et al. 2014; Morrison et al. 2015). Special genome editing tools were created, such as a negative selection cassette and a self-deleting cassette to improve transformation efficiency while reducing the amount of genome perturbation during strain creation (Sung et al. 2001; Weng et al. 2009). It is clear from the above studies that high efficiencies of transformation and marker-less genome editing can be achieved with large PCR amplicon donors.

Despite the extensive measurements of parameters affecting transformation efficiency in *S. pneumoniae*, there has not been a measure of the dependence of transformation efficiency on genomic versus PCR amplicon donors in the commonly used laboratory strain, Rx. To investigate the dependence of transformation efficiency on amplicon homology to the recipient, I compared single base changes to 1.2-kb (kilobase) cassettes with flanking homologous arms. Increasing the amount of homologous regions in the donor increased transformation efficiency for PCR amplicons, and transformation efficiency was ~100-fold higher for PCR amplicons compared to

genomic DNA for every amount of donor DNA tested. The high efficiency of recombination during natural transformation of *S. pneumoniae* measured here allows for markerless and selection-less genome editing, as wells as creation of multiple mutants of an essential gene without the disruption of neighboring regions.

5.3 Results

5.3.1 Transformation efficiency depends on the amount of donor and the topology of the donor.

In comparison of genomic DNA donors with a 7-kb PCR amplicon carrying the same point mutation, there is ~300-fold more complexity in the genomic donor. A DNA preparation from a genomic donor presents a challenge to efficient genome editing because the majority of DNA is extraneous to the intended gene replacement. To achieve a high efficiency of genome modification, a PCR-amplified donor containing only the allelic replacement of interest and flanking homology was tested for an increase in transformation efficiency compared to the genomic donor. To investigate whether the decrease in complexity of donor results in an increase in transformation efficiency, I compared transformation efficiency for the two different types of donors (Fig. 23). I also compared the transformation efficiency for four different PCR amplicon donors. Two of the PCR amplicon donors had single-base substitutions, and two had ~1-kb cassettes flanked by homologous arms.

To test the dependence of transformation efficiency on complexity of the DNA donor, I compared genomic or a 7.4-kb PCR amplified donor (with a single-base substitution), I monitored the amount of transformants over a 100 million-fold range of the amount of DNA donor (Fig. 24). As expected, the PCR amplicon donor transformed at a rate ~300-fold above the genomic donor at every amount of DNA in the linear range. Transformants appeared to be saturating at 100 ng/mL

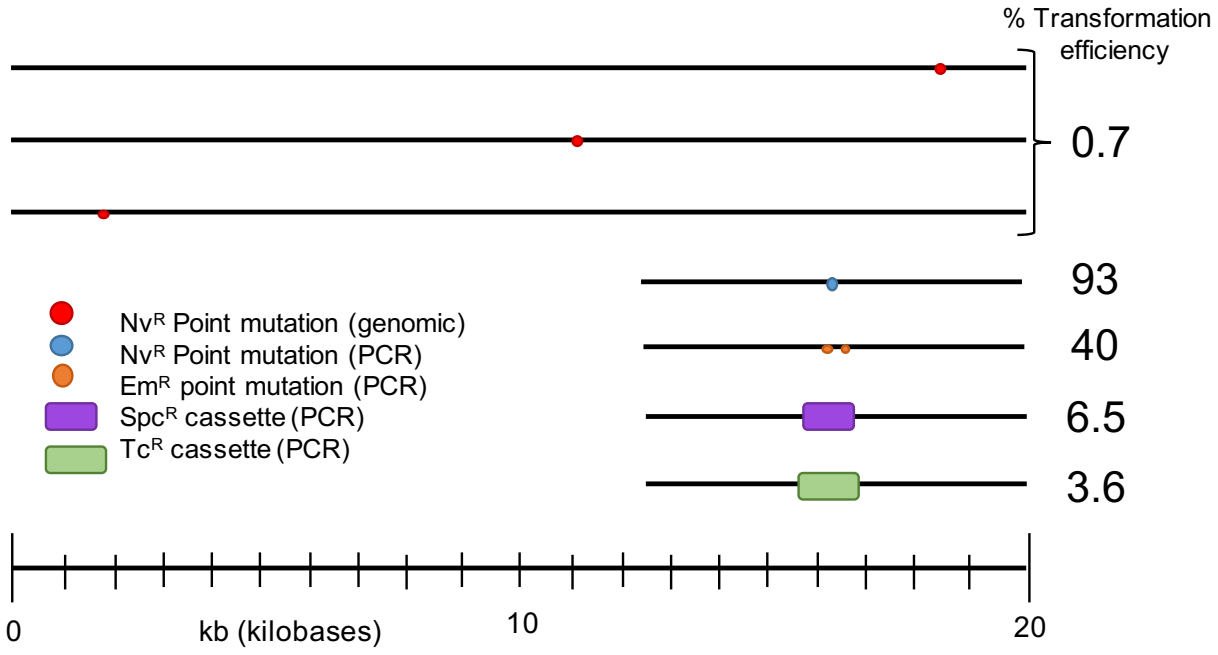


Figure 23. High efficiency gene replacement with multiple donors. Schematic of different donors used and transformation efficiency observed at 100 ng/mL of donor DNA. Genomic donor with a one-base substitution is shown by black lines representing homology and one spot representing the substitution. Sheared genomic donors are represented as 20 kb fragments. Also tested were 7.4-kb PCR amplicon donors with four differences: (1) a one base substitution (100% homology to recipient), (2) two base substitutions (100% homology to recipient, (3) a 1-kb cassette (86% homology to recipient), (4) a 1.2- cassette (83% homology to recipient). On the left, a chart displays the type of donor, the approximate % homology of the marker to the recipient, and the % transformants relative to cfu seen at 100 ng/mL. Black lines, DNA; red dots, Nv^R point mutation in the genomic donor; blue dots, Nv^R point mutation in the PCR donor, orange, point mutations in the PCR donor; purple box, Spc^R cassette with no homology to recipient; green, 1.2-kb cassette with no homology to the recipient; kb ruler, each hash mark is 1 kb.

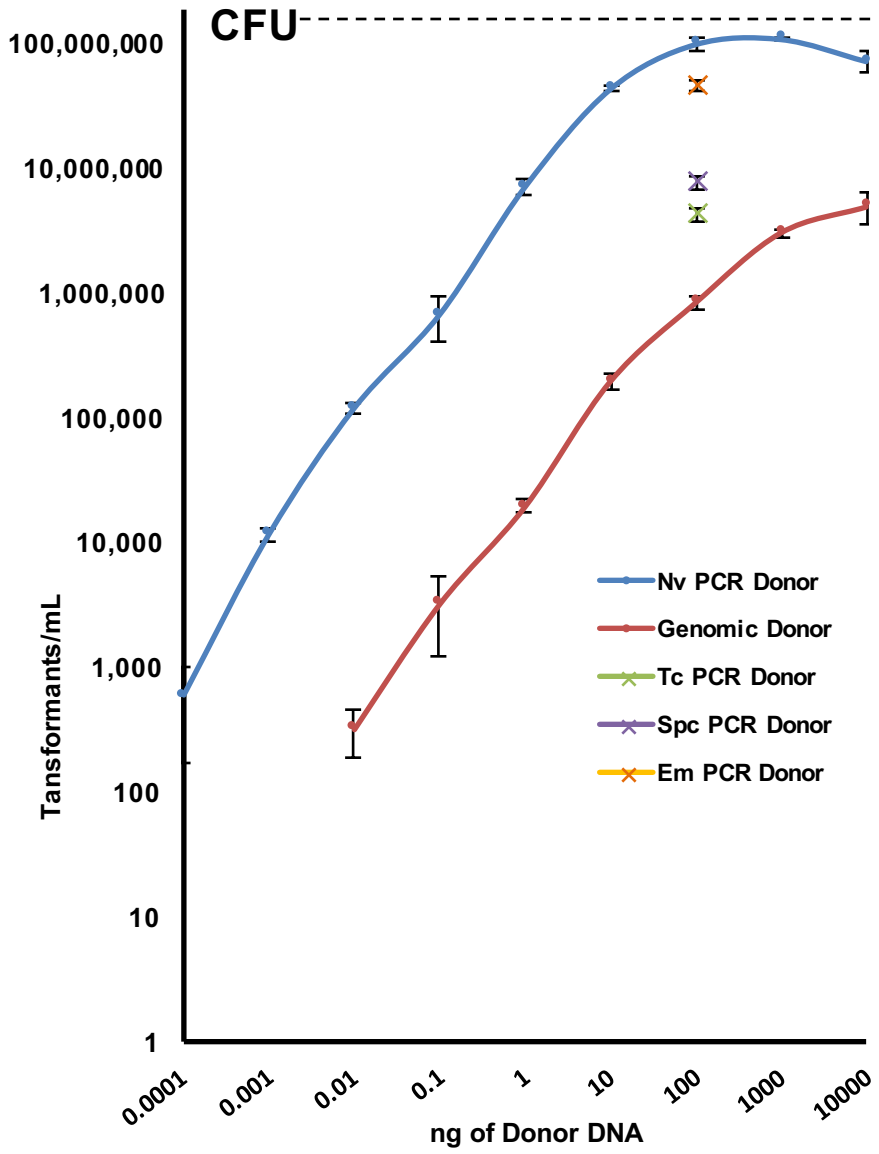


Figure 24. Transformation efficiency dependence on the amount of donor DNA. The transformation efficiency was measured by 10-fold increments over a range of amounts from 0.0001 ng/mL to 10,000 ng/mL of donor, for both a genomic donor and a 7.4-kb PCR amplicon. The transformation efficiency of several other PCR amplicons was measured at 100 ng/mL of donor DNA. Black hashed line, cfu; blue line, transformants/mL for a 7.4-kb PCR amplicon donor differing by one base; red line, transformants/mL for genomic donor differing by one base; orange X, transformation efficiency of a 7.4-kb marker with two base substitutions; purple X, transformation efficiency of a 7.4-kb marker with a 1-kb cassette; green X, transformation efficiency of a 7.4-kb marker with a 1.2-kb cassette.

with PCR donor and 10,000 ng/mL with genomic donor. To compare different constructs of PCR amplified donor DNA, I compared transformation efficiency with four different 7-kb PCR donors: (1) a one base difference, (2) a two base difference, (3) a 1-kb cassette, (4) a 1.2-kb cassette (Fig. 23, 24). As expected, I observed a sharp decrease in transformation efficiency with the decrease in percent homology of the donor.

5.3.2 Markerless genome editing without the use of selective resistance markers

Nearly 100% transformation efficiency was achieved from a 7.4-kb donor that differed from the recipient by one base (Fig. 23, 24). This donor conferred antibiotic resistance to Novobiocin by a single base change resulting in a serine to leucine change at amino acid 127 in *gyrB* (Muñoz et al. 1995). The near-100% transformation efficiency of this donor indicated that it is possible to achieve high efficiency integration of markerless donors with small changes such as single–base substitutions. To investigate whether markerless changes can be made using PCR amplicons, the high transformation efficiency was used to edit essential genes without selectable markers or any extraneous DNA. To test essential gene editing with this method, four independent PCR amplicons of mutant *rpoD* sequences were transformed into the WT. Each amplicon varied from the recipient by one base. *rpoD* encodes the primary σ factor and gene disruptions with selectable markers would lead to death. After transformation and non-selective plating, 10 to 20 colonies were sampled for each cross. 60% of total colonies tested took up and incorporated the mutant donor DNA (Table XI). The confirmation of mutant or WT alleles after transformation was tested by using the mismatched PCR assay (Fig. 25), and is also described in Methods and Chapter 3. Sanger sequencing of positive subclones confirmed integration of the single-base change, and also identified the lack of incorporation of any mutations in the sequence other than the desired

Table X. Markerless, selection-less replacement of an essential gene occurs at high efficiency with a 7.4-kb PCR amplicon donor.

Donor[a]	Colonies Tested[b]	Sequencing Result[c]			
		WT[d]	Mutant[e]	Mixed[f]	No Result[g]
NYT1 *rpoD*	20	5	13	1	1
ALT4 *rpoD*	10	4	5	0	1
FLT4 *rpoD*	10	4	6	0	0
ILT1 *rpoD*	10	4	5	0	1
Total	50	17	29	1	3

[a]7.4-kb PCR amplified markerless DNA donor

[b]Amount of colonies surveyed for transformation

[c]Sequence as determined by mismatch PCR

[d]WT allele at this position

[e]Mutant allele at this position

[f]Both mutant and WT alleles identified

[g]Failed PCR

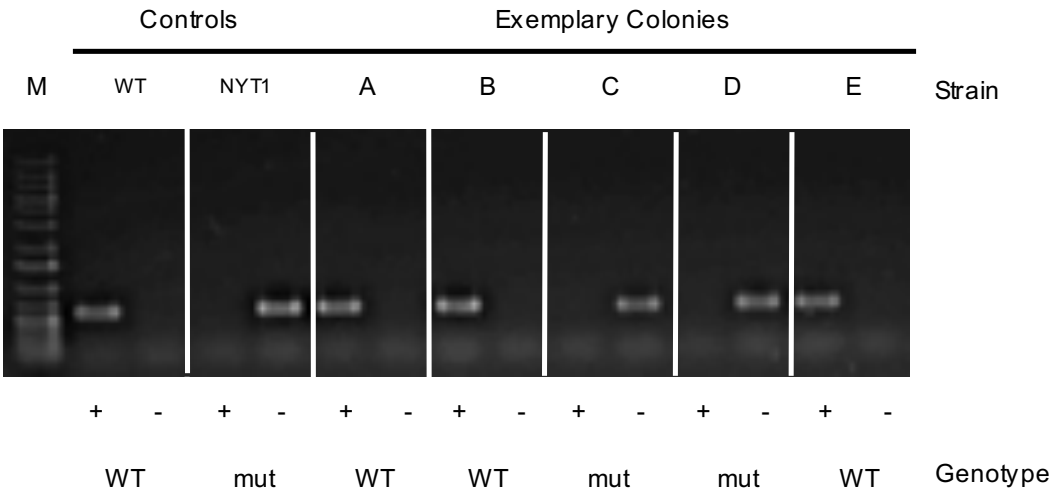


Figure 25. Gel analysis of PCR products from the mismatched PCR assay. Exemplary colonies were examined for transfer of the *rpoD* single-base substitutions into the WT strain. +, primer pair specific for WT sequence; -, primer pair specific for mutant sequence. M, molecular weight standard; WT, parent strain; NYT1, suppressor strain; A-E, colonies tested; genotype, identified allele at single base.

change. High efficiency genome editing of *S. pneumoniae* at a single locus was achieved without the use of selectable markers or gene disruption.

5.4 Discussion

High efficiency genome editing that disrupts as little of the genome as possible is a valuable tool for strain creation for genetic study. In the *S. pneumoniae* laboratory strain Rx, high efficiency editing was achieved with large donors delivered in near-saturating amounts. Applying this method to other genes will ensure accurate and efficient transformation to quickly create new strains. These new strains can easily be tested for correct integration using mismatched-primer PCR.

The strategy explored here for Rx strains, derivatives of the R6 strain of *S. pneumoniae*, reveals that markerless, scar-less, and selection-less genome editing can be achieved with high efficiency in strains of *S. pneumoniae* that are highly transformable. However, many strains of *S. pneumoniae*, such as encapsulated clinical isolates like D39, do not have as high transformation efficiency as reported for this Rx strain (Yother et al. 1986). Methods to perform scar-less and selection less genome editing in low transformability strains have been developed (Dalia et al. 2014), where using co-transformation with a selectable marker to isolate transformable cells simultaneously with a second selection-less donor led to edits in over ~50% of isolates. Using this method, selection-less, markerless, and scar-less genome editing can be achieved if strains have low frequencies of competent cells. Using either method, the single-locus method described here for highly transformable strains or that described in Dalia 2014 for low transformability strains, one can achieve efficient and rapid genome editing at many loci to create new strains for genetic study.

VI. GENERAL CONCLUSION AND DISCUSSION

Before the start of this work, several conclusions had been made about the function of ComW, but its exact role during transformation in *S. pneumoniae* remained unclear. ComW is required for the full transformation response (Bartilson et al. 2001), σ^X activity, and σ^X stability (Luo et al. 2004; Sung & Morrison 2005). To elucidate the role of ComW during transformation, I used a suppressor screen to retrieve rare clones that bypass the ComW requirement for transformation. I identified several single-base substitutions in σ^A, the primary σ factor, that relieve the need for ComW (Tovpeko & Morrison 2014). Eight of the ten bypass residues mapped to residues implicated in affinity and binding of core RNAP (Nickels et al. 2005; Dove et al. 2003; Laurie et al. 2003). Late gene expression is also restored in σ^A* mutants. Together, these data suggest that ComW's role is to act as a σ factor activator, to promote σ^X binding to core RNAP at the point of σ switching from σ^A to σ^X during transformation.

Based on previously studied mechanisms of alternative σ factor regulation in other bacteria (Österberg et al. 2011), I made several predictions for the identity of the suppressor (Fig. 26). For example, the protein RsiW sequesters σ^W in *B. subtilis*, and must be cleaved by YluC to release σ^W to bind core RNAP (Schöbel et al. 2004). If ComW's role is to promote σ^X activity by releasing it from sequestration, I would expect a suppressor mutation in a protein similar to the sequestering protein RsiW to restore transformation in *comW* mutants (Fig. 26). In *S. thermophilus*, σ^X is negatively regulated by degradation through the MecA adaptor protein and the Clp protease complex (Boutry et al. 2012). If ComW's role is to promote σ^X activity by preventing degradation by a similar complex, I would expect suppressor mutations in proteases that may be cleaving σ^X (Fig. 1). In *E. coli*, the small protein Crl directly binds σ^S and promotes σ^S binding to core RNAP (Gaal et al. 2006). If ComW's role is to promote σ^X binding to core RNAP, I would expect

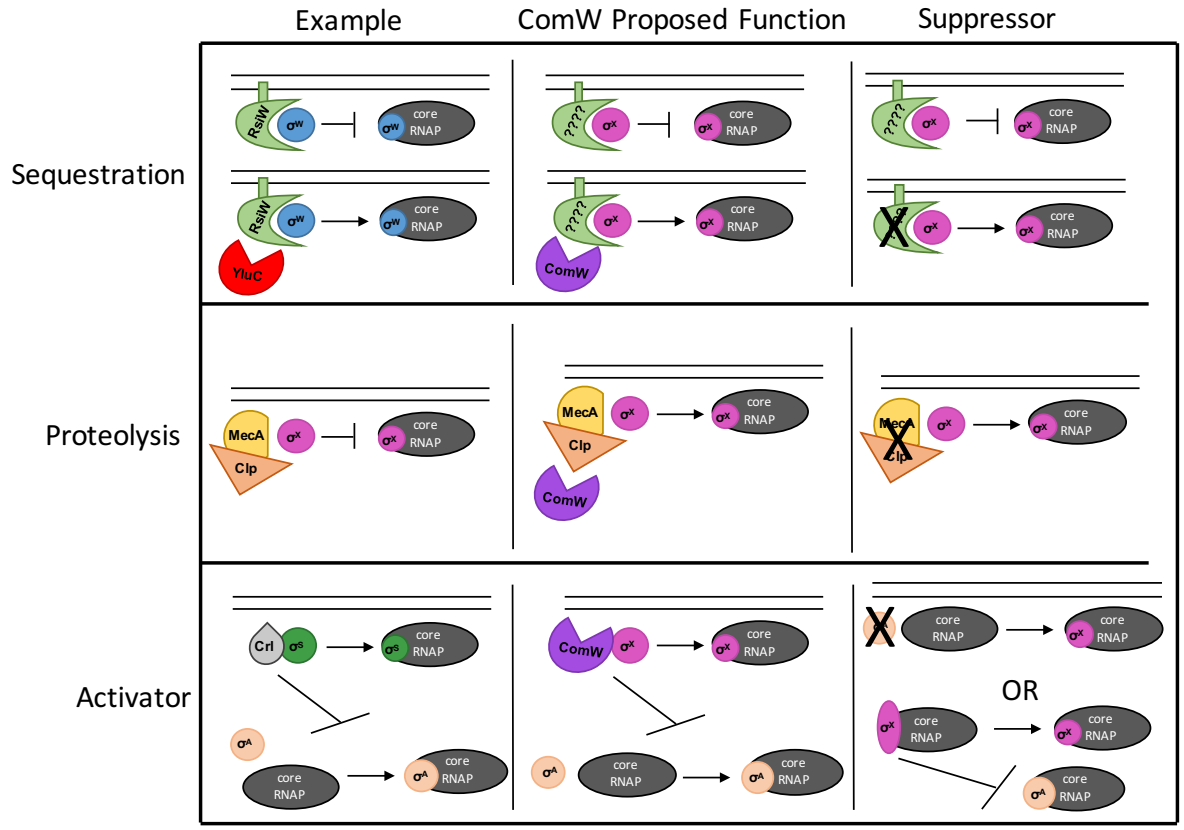


Figure 26. Examples of alternative σ factor regulation and predictions of possible functions for suppressors of ComW. Three examples of alternative σ factor regulation; sequestration, proteolysis, and activators are shown from top to bottom in the left column. Possible functions for ComW in the context of the proposed method of regulation are shown in the middle column. Possible suppressors in *comW* mutants are shown in the right column. Top row: in *B. subtilis*, RsiW sequesters σ^W, RsiW is cleaved by YluC releasing σ^W (left). ComW's releases σ^X from sequestration (middle), a suppressor mutation in a sequestering protein restores core binding (right). Middle row: in *S. thermophilus*, σ^X is negatively regulated by degradation (left). ComW prevents degradation (middle), and a suppressor mutation in a protease restores core binding (right). Bottom row: in *E. coli*, Crl binds σ^S to promote binding to core RNAP, preventing σ^A binding to core (left). ComW functions as Crl (middle), and a suppressor mutation in σ^A or σ^X restores σ^X binding to core (right). Black double lines, cell wall/membrane; green moon, RsiW; blue circle, σ^W; gray oval, core RNAP; red, YluC; purple, ComW; magenta circle, σ^X; yellow half-circle, MecA; orange triangle, Clp protease complex; dark green, σ^S; gray teardrop, Crl; X, suppressor mutation; magenta oval, mutated σ^X; beige circle, σ^A.

mutations in the primary σ factor that reduced primary σ factor affinity for core RNAP or mutations in σ^X that increased σ^X affinity for core RNAP (Fig. 26). Interestingly, in my screen, all suppressor mutations were in *rpoD*, the gene encoding the primary σ factor, pointing to ComW's role in facilitating σ^X assembly with core RNAP.

Single-base substitutions in σ^A bypassed the need for ComW during transformation, and the co-location of these residues with those implicated in σ factor affinity for core RNAP immediately implicated ComW in promoting σ^X binding to core RNAP. However, if this is ComW's role, it is interesting that the suppressor screen did not also yield mutations in σ^X that improve σ^X binding to core RNAP. It is possible that such mutations did arise, but were outcompeted by the σ^A mutations during the enrichment. Also, since *S. pneumoniae* has two copies of *comX*, encoding σ^X, it is possible that a suppressor mutation in one copy of *comX* was recessive to the WT copy of *comX*, and such mutations were not collected at the end of the enrichment. Regardless, it would be interesting to test σ^X mutants for binding to core RNAP, potentially identifying resides that are important for affinity to core RNAP and residues that may be binding ComW. Using error-prone PCR of *comX* and testing this pool of mutated *comX* for transformation restoration in *comW* mutants is likely to reveal specific residues or regions of σ^X that are important for binding core RNAP.

In ComW mutants, the amount of transformants is 10^4-fold below the WT level, the amount of late gene expression is 10% of the WT level, and the amount of σ^X is 10% of the WT level (Sung & Morrison 2005; Luo et al. 2004). In this study, I observed partial restoration of transformation and late gene expression in Δ*comW* σ^A* mutants. However, this does not exclude the possibility that the amount of σ^X present is also restored. During transformation, σ^X amounts peak at 15 minutes after CSP induction and then rapidly decrease. To determine whether the amount of σ^X is

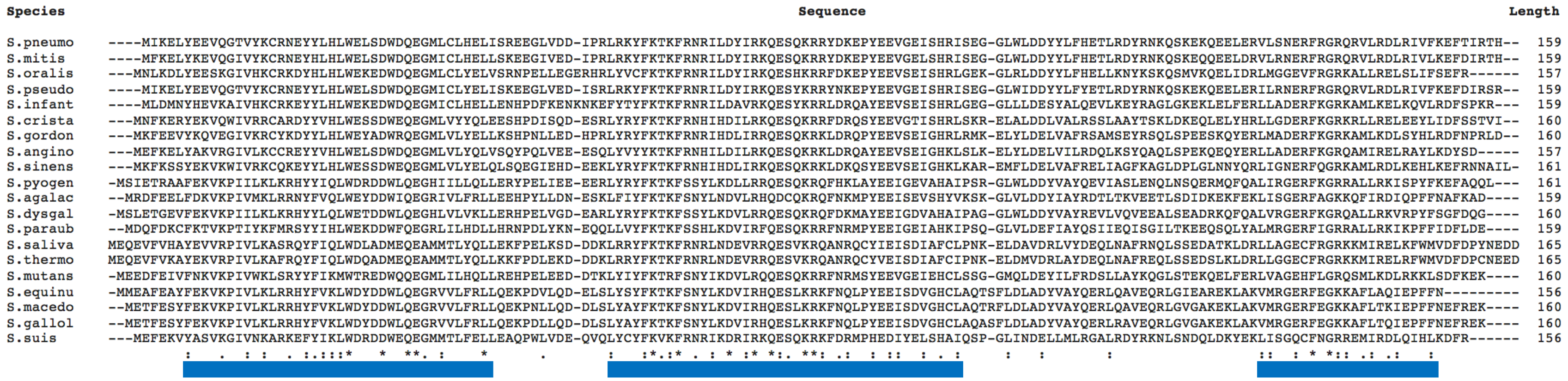
partially, or fully, restored in Δ*comW* σ^A* strains, we need to measure the amount of σ^X present by quantitative Western Blotting over a period of 30 minutes after the induction of CSP. Comparing these levels to a purified σ^X standard will show whether the mechanism of transformation restoration in σ^A* mutants is to restore the amount of σ^X levels as well as to restore late gene expression levels.

The differences in distribution of *comX* and *comW* sequences amongst the streptococci prompts several questions about ComW's regulation of σ^X activity during transformation. At least one copy of *comX* has been identified in every single sequenced species of streptococcus, with most species having two identical copies in their genome (Martin et al. 2006; Johnston, Martin, et al. 2014). Many of the sequenced species are transformable, such as *S. mutans* in the mutans group of streptococci (Li et al. 2001; Mashburn-Warren et al. 2010), *S. thermophilus* and *S. salivarius* in the salivarius group (Fontaine et al. 2010), *S. infantarius* and *S. macedonicus* in the bovis group (Morrison et al. 2013), and *S. pyogenes* in the pyogenic group (Marks et al. 2014). However, ComW has only been identified in eleven species of streptococci limited to the mitis and anginosus groups. In *S. pneumoniae*, the full σ^X response relies on ComW, but what is the mechanism of σ^X regulation in the species that do not have ComW? Why are some versions of σ^X able to incite transformation without the assistance of ComW in some streptococci but not others? Further study into the σ^X regulatory cascades in species that do not have ComW is necessary to understand the role of ComW in *S. pneumoniae* and how other streptococci achieve transformation without it.

Data from this study points to ComW acting as a σ switching factor that assists in the change from σ^A to σ^X during transformation. Species that do not have ComW may overcome the need for ComW during transformation in two ways, either by having a σ^A with lower affinity for core RNAP compared to the *S. pneumoniae* σ^A or by having a σ^X with higher affinity for core

RNAP compared to the *S. pneumoniae* σ^X. In this study, I compared the σ^A from *S. mutans*, a naturally transformable species of streptococci that does not have ComW, to the σ^A from *S. pneumoniae*, and found that the *S. mutans* σ^A is able to partially bypass the ComW requirement for transformation. This suggests that streptococcus species without ComW may be overcoming the need for ComW displayed by *S. pneumoniae* by having σ^A proteins with lower affinities for core RNAP. However, it is also possible that the species without ComW have σ^X proteins that have higher affinities for core RNAP compared to that of *S. pneumoniae* σ^X. An alignment of σ^X sequences from over twenty species shows regions of homology, but the resides that differ amongst the species may be the source of species differences in σ^X affinity for core RNAP (Fig. 27). To test whether σ^X from different species can bypass the need for ComW in *S. pneumoniae*, it would be interesting to replace the pneumococcal σ^X with sequences from other streptococci and test for transformation. A complementary approach would be to also replace the σ^X sequences of transformable streptococci with the *S. pneumoniae* σ^X and test for transformability. These studies may reveal more information about ComW's role in σ^X regulation during transformation.

ComW itself is an interesting protein for several reasons. First, the only proteins bearing any homology to ComW are ComW sequences from other species of streptococci, and these sequences are found only in the mitis and anginosus groups. In the various strains of *S. pneumoniae*, the ComW amino acid identity ranges from 100% to 76%, with long regions of highly conserved residues (Fig. 28). *S. pneumoniae's* closest relative, *S. mitis,* is a diverse species that contains more difference amongst the strains than there are differences between any two species of streptococci (Bek-Thomsen et al. 2008; Kilian et al. 2014). This variety in strains and in sequences of *S. mitis* is similarly reflected in the range of ComW sequences from various *S. mitis*



```
Species                                                                    Sequence                                                                                 Length

S.pneumo      ----MIKELYEEVQGTVYKCRNEYYLHLWELSDWDQEGMLCLHELISREEGLVDD-IPRLRKYFKTKFRNRILDYIRKQESQKRRYDKEPYEEVGEISHRISEG-GLWLDDYYLFHETLRDYRNKQSKEKQEELERVLSNERFRGRQRVLRDLRIVFKEFTIRTH--  159
S.mitis       ----MFKELYKEVQGIVYKCRNEYHLHLWELSDWDQEGMICLHELLSKEEGIVED-IPRLRKYFKTKFRNRILDYIRKQESQKRRYDKEPYEEVGELSHRISEG-GLWLDDYYLFHETLRDYRNKQSKEQQEELDRVLRNERFRGRQRVLRDLRIVLKEFDIRTH--  159
S.oralis      ---MNLKDLYEESKGIVHKCRKDYHLHLWEKEDWDQEGMLCLYELVSRNPELLEGERHRLYVCFKTKFRNRILDYIRKQESHKRRFDKEPYEEVSEISHRLGEK-GLRLDDYYLFHELLKNYKSKQSMVKQELIDRLMGGEVFRGRKALLRELSLIFSEFR------  157
S.pseudo      ----MIKELYEEVQGTVYKCRNEYYLHLWELSDWDQEGMICLYELISKEEGLVED-ISRLRKYFKTKFRNRILDYIRKQESYKRRYNKEPYEEVGEISHRISEG-GLWIDDYYLFYETLRDYRNKQSKEKQEELERILRNERFRGRQRVLRDLRIVFKEFDIRSR--  159
S.infant      ----MLDMNYHEVKAIVHKCRKEYYLHLWEKEDWDQEGMICLHELLENHPDFKENKNKEFYTYFKTKFRNRILDAVRKQESYKRRLDRQAYEEVSEISHRLGEG-GLLLDESYALQEVLKEYRAGLGKEKLELFERLLADERFKGRKAMLKELKQVLRDFSPKR---  159
S.crista      ---MNFKERYEKVQWIVRRCARDYYVHLWESSDWEQEGMLVYYQLEESHPDISQD-ESRLYRYFKTKFRNHIHDILRKQESQKRRFDRQSYEEVGTISHRLSKR-ELALDDLVALRSSLAAYTSKLDKEQLELYHRLLGDERFKGRKRLLRELEEYLIDFSSTVI--  160
S.gordon      ---MKFEEVYKQVEGIVKRCYKDYYLHLWEYADWRQEGMLVLYELLKSHPNLLED-HPRLYRYFKTKFRNRIHDLIRRQESQKRKLDRQPYEEVSEIGHRLRMK-ELYLDELVAFRSAMSEYRSQLSPEESKQYERLMADERFKGRKAMLKDLSYHLRDFNPRLD--  160
S.angino      ---MEFKELYAKVRGIVLKCCREYYVHLWELSDWDQEGMLVLYQLVSQYPQLVEE-ESQLYVYYKTKFRNHILDILRKQESQKRKLDRQAYEEVSEIGHKLSLK-ELYLDELVILRDQLKSYQAQLSPEKQEQYERLLADERFKGRQAMIRELRAYLKDYSD-----  157
S.sinens      ---MKFKSSYEKVKWIVRKCQKEYYLHLWESSDWEQEGMLVLYELQLSQEGIEHD-EEKLYRYFKTKFRNHIHDLIRKQESQKRKLDKQSYEEVSEIGHKLKAR-EMFLDELVAFRELIAGFKAGLDPLGLNNYQRLIGNERFQGRKAMLRDLKEHLKEFRNNAIL-  161
S.pyogen      -MSIETRAAFEKVKPIILKLKRHYYIQLWDRDDWLQEGHIILLQLLERYPELIEE-EERLYRYFKTKFSSYLKDLLRRQESQKRQFHKLAYEEIGEVAHAIPSR-GLWLDDYVAYQEVIASLENQLNSQERMQFQALIRGERFKGRRALLRKISPYFKEFAQQL---  161
S.agalac      --MRDFEELFDKVKPIVMKLRRNYFVQLWEYDDWIQEGRIVLFRLLEEHPYLLDN-ESKLFIYFKTKFSNYLNDVLRHQDCQKRQFNKMPYEEISEVSHYVKSK-GLVLDDYIAYRDTLTKVEETLSDIDKEKFEKLISGERFAGKKQFIRDIQPFFNAFKAD----  159
S.dysgal      -MSLETGEVFEKVKPIILKLKRHYYLQLWETDDWLQEGHLVLVKLLERHPELVGD-EARLYRYFKTKFSSYLKDVLRRQESQKRQFDKMAYEEIGDVAHAIPAG-GLWLDDYVAYREVLVQVEEALSEADRKQFQALVRGERFKGRQALLRKVRPYFSGFDQG----  160
S.paraub      --MDQFDKCFKTVKPTIYKFMRSYYIHLWEKDDWFQEGRLILHDLLHRNPDLYKN-EQQLLVYFKTKFSSHLKDVIRFQESQKRRFNRMPYEEIGEIAHKIPSQ-GLVLDEFIAYQSIIEQISGILTKEEQSQLYALMRGERFIGRRALLRKIKPFFIDFLDE----  159
S.saliva      MEQEVFVHAYEVVRPIVLKASRQYFIQLWDLADMEQEAMMTLYQLLEKFPELKSD-DDKLRRYFKTKFRNRLNDEVRRQESVKRQANRQCYIEISDIAFCLPNK-ELDAVDRLVYDEQLNAFRNQLSSEDATKLDRLLAGECFRGRKKMIRELKFWMVDFDPYNEDD  165
S.thermo      MEQEVFVKAYEKVRPIVLKAFRQYFIQLWDQADMEQEAMMTLYQLLKKFPDLEKD-DDKLRRYFKTKFRNRLNDEVRRQESVKRQANRQCYVEISDIAFCIPNK-ELDMVDRLAYDEQLNAFREQLSSEDSLKLDRLLGGECFRGRKKMIRELRFWMVDFDPCNEED  165
S.mutans      -MEEDFEIVFNKVKPIVWKLSRYYFIKMWTREDWQQEGMLILHQLLREHPELEED-DTKLYIYFKTRFSNYIKDVLRQQESQKRRFNRMSYEEVGEIEHCLSSG-GMQLDEYILFRDSLLAYKQGLSTEKQELFERLVAGEHFLGRQSMLKDLRKKLSDFKEK----  160
S.equinu      -MMEAFEAYFEKVKPIVLKLRRHYFVKLWDYDDWLQEGRVVLFRLLQEKPDVLQD-ELSLYSYFKTKFSNYLKDVIRHQESLKRKFNQLPYEEISDVGHCLAQTSFLDLADYVAYQERLQAVEQRLGIEAREKLAKVMRGERFEGKKAFLAQIEPFFN---------  156
S.macedo      --METFESYFEKVKPIVLKLRRHYFVKLWDYDDWLQEGRVVLFRLLQEKPNLLQD-DLSLYAYFKTKFSNYLKDVIRHQESLKRKFNQLPYEEISDVGHCLAQTRFLDLADYVAYQERLQAVEQRLGVGAKEKLAKVMRGERFEGKKAFLTKIEPFFNEFREK----  160
S.gallol      --METFESYFEKVKPIVLKLRRHYFVKLWDYDDWLQEGRVVLFRLLQEKPDLLQD-DLSLYAYFKTKFSNYLKDVIRHQESLKRKFNQLPYEEISDVGHCLAQASFLDLADYVAYQERLRAVEQRLGVGAKEKLAKVMRGERFEGKKAFLTQIEPFFNEFREK----  160
S.suis        ---MEFEKVYASVKGIVNKARKEFYIKLWDRDDWEQEGMMTLFELLEAQPWLVDE-QVQLYCYFKVKFRNRIKDRIRKQESQKRKFDRMPHEDIYELSHAIQSP-GLINDELLMLRGALRDYRKNLSNDQLDKYEKLISGQCFNGRREMIRDLQIHLKDFR------  156
                       :   .  : :  . :.:::*   *  **. :    *      .       :   :*.:* . : * :* *:. **: .:  : ::  : . :     :   :       :                 ::  : * *:: .: .:   :
```

Figure 27. Alignment of σ^X homologs from *S. pneumoniae* and other streptococcal species. *S. pneumoniae* σ^X residues (accession number AAK98817.1) were aligned with corresponding residues of σ^X sequences from *S. mitis* (KEQ34456.1), *S. oralis* (EFU63939.1), *S. pseudopneumoniae* (EID29564.1), *S. infantis* (EFX36175.1), *S. cristatus* (EGU68770.1), *S. tigurinus* (WP_007522820.1), *S. gordonii* (KJQ56223.1), *S. anginosus* (AGU82552.1), *S. sinensis* (KGM37423.1), *S. pyogenes* (ACI60595.1), *S. agalactiae* (CCO74071.1), *S. dysgalactiae* (BAN92743.1), *S. parauberis* (EGE55156.1), *S. salivarius* (CCB94501.1), *S. thermophilus* (ADQ62208.1), *S. mutans* (NP_722295.1), *S. equinus* (EFM26373.1), *S. macedonicus* (WP_044563186.1), *S. gallolyticus* (CBI12727.1), and *S. suis* (CAR77110.1). Alignment was made using Clustal Omega with default parameters (Sievers et al. 2011). Species name is on left, amino acid sequence is in middle, protein length is on right. Blue bars, tracts of conserved resides that may be regions of conserved function; black text, regions of non-conservation and potential sources for differences in affinity for core RNA polymerase. Asterisks, identical residues; colons, conserved residues; periods, semi-conserved residues.

```
Strain                                                                                                  % Identity
                                                                                                          to R6
R6            MIMLQKIYEQMANFYDSIEEEYGPTFGDNFDWEHVHFKFLIYYLVRYGIGCRKDFIVYHYRVAYRLYLEKLVMNRGFISC          --
430772        --MLQKIYEQMANFYDSIEEEYGPTFGDNFDWEHVHFKFLIYYLVRYGIGCRKDFIIYHYRVAYRLYLEKLVMNRGFISC          99
2082239       --MLQKIYEQMANFYDSIEEEYGPTFGDNFDWEHIHFKFLIYYLVRYGIGCRRDFIVYHYRVAYRLYLEKLVMNRGFISC          97
SMRU1423      --MLQKIYEQMANFYDSIEEEYGPTFGDNFDWEYVHFKFLIYYLVRYGIGCHRDFIVYHYRVAYRLYLEKLVMNRGFISC          96
SMRU577       --MLKSIYEQMANFYDSIEEEYGPTFGDNFDWEHVHFKFLIYYLVRYGIGCRRDFIVYHYRVAYRLYLEKMIMKQGFISC          91
LMG1278       --MLQKIYEQMADFYRNIEEEYGTVFGDNFDWEHFHFKFLIYYLVRYGIGCRRDFIVYHYRVAYRLYLEKMIMNRGFISC          88
SMRU737       --MLQKIYEQMADFYRNIEEEYGTVFGDNFDWEHFHFKFLIYYLVRYGIGCRRDFIVYHYRVAYRLYLEKMIMKQGFISC          86
SMRU2014      --MLQSIYDQMTDFYDSIEEEYATFFGNSWDWEHFHFKFLIYYLVRYGIGCRRDFIVYHYRVTYRLYLEKMIMNRGFISC          81
19F           --MLQNIYDQMTDFYDSIEEEYATFFDNSWNWEHFHFKFLIYYLVRYGIGCRRDFIVYHYRVAYRLYLEKMIMNRGFISC          79
SMRU946       --MLQNIYDQMTDFYDSIEEEYATFFDNSWEWEHFHFKFLIYYLVRYGIGCRRDFIVYHYRVAYRLYLEKMIMKQGFISC          77
SMRU824       --MLQSIYDQMTDFYDSIEEEYATFFDNSWEWEDFHFKFLIYYLVRYGIGCRRDFIVYHYRVAYRLYLEKMIMKQGFISC          76
                **:.**:**::** .*****.  * :.::** .****************::***:*****:*******::*::*****
```

Figure 28. Alignment of ComW homologs from various strains in *S. pneumoniae.* *S. pneumoniae* strain R6 ComW residues (accession number AAK98824.1) were aligned with corresponding residues of ComW sequences from *S. pneumoniae* strains 430772 (WP_05439300.1), 2082239 (EJG75956.1), SMRU1423 (COP90535.1), SMRU577 (COB21985.1), LMG1278 (CNA73777.1), SMRU737 (CEY67237.1), SMRU2014 (CJD59153.1), 19F (CEY35379.1), SMRU946 (CIY80391.1), and SMRU824 (CON90623.1). Alignment was made using Clustal Omega with default parameter (Sievers et al. 2011). Red text, regions of conservation; black text, regions of residues that are non-conserved. Strain name is on the left, sequence is in the middle, percent identity to R6 sequence on right. Asterisks, identical residues; colons, conserved residues; periods, semi-conserved residues.

strains that range in identity from 88% to 42% to the *S. pneumoniae* sequence (Fig. 29). Some strains of *S. pneumoniae* and *S. mitis* have identical ComW protein sequences, such as the sequences from *S. pneumoniae* LMG1278 and *S. mitis* SK579, suggesting that DNA exchange is occurring between these two species in the environment. It would be interesting to perform ComW allele replacements from species to species, by transforming several of the low identity sequences into *S. pneumoniae* and observing their effect on transformation efficiency to see whether function is dependent on the amount of homology of ComW to the *S. pneumoniae* sequence, or whether the conserved residues will retain function regardless of the adjacent non-conserved residues. An alignment of all available ComW sequences reveals two sets of conserved residues that may be possible motifs important to ComW function (Fig. 30). The first is ExEY and the second is LLYYL (Fig. 30, bold). However, if the regions of conservation are extended to include residues that are semi-conserved, there appear to be six regions that are similar in all sequences of ComW (Fig. 30, purple). These conserved regions are likely key to ComW's function, and their topology in the 3D structure of ComW will reveal clues towards ComW's role during transformation.

The location of the conserved residues in the 3D structure of ComW may reveal clues about the function of the conserved residues. To aid in elucidating the function of ComW, we can employ structural modeling to predict the 3D structure of ComW. Many structure prediction programs, such as Phyre2 (Kelley et al. 2015), are based on modeling proteins on structures that already exist. This modeling system predicts that ComW has three alpha helices that are loosely packed (Fig. 31A) However, since ComW is a unique protein with no homology to previously modeled proteins, we would need to employ a structural modeling program that can predict a structure *ab initio* from ComW sequences from various species to ensure a more accurate prediction of ComW structure. An example of an *ab initio* structural modeling program that can predict ComW's structure is the

```
Species/Strain                                          Sequence                                   % Identity
                                                                                                    to R6
R6                 MIMLQKIYEQMANFYDSIEEEYGPTFGDNFDWEHVHFKFLIYYLVRYGIGCRKDFIVYHYRVAYRLYLEKLVMNRGFISC   --
SK579              --MLQKIYEQMADFYRNIEEEYGTVFGDNFDWEHFHFKFLIYYLVRYGIGCRRDFIVYHYRVAYRLYLEKMIMNRGFISC   88
SK616              --MLQKIYEQMADFYRNIEEEYGTVFGDNFDWEHFHFKFLIYYLVRYGIGCRRDFIVYHYRVAYRLYLEKMIMKQGFISC   86
SK1073             --MLQSIYDQMTDFYDSLEEEYATFFGDSWDWEHFHFKFLIYYLVRYGIGCRRDFIVYHYRVAYRLYLEKLIMNQGFISC   82
SK597              --MLQNIYDQMTDFYDSIEEEYATFFGNSWDWEHFHFKFLIYYLVRYGIGCRRDFIVYHYRVAYRLYLEKMIMKQGFISC   79
SK1080             --MLQSIYDQMTDFYDSIEEEYAIFFGNSWDWEHFHFKFLIYYLVRYGIGCRRDFIVYHYRVAYRLYLEKLIMKQGFVAC   78
SK1126             --MLQSIYDQMTDFYDSIEEEYATFFGDSWDWEHFHFKFLIYYLFRYSIGNHRDFIVYHYRVAYRLYLEKMIMKQGFISC   76
SK564              MLTLQNIYDQMTDFYDSIEEEYATFFDNSWEWEHFHFKFLIYYLVRYGIGCRRDFIVYHYRVAYRLYLEKLIMKQGFVAC   74
11/5               --MLQSIYNQMTDFYDSIEEEYATFFDNSWEWEHFHFKFLIYYLVQYDICRRRDFIVYHYRVAYRLYLEKMIMKQGFISC   72
bv.2str.F0392      --MFQDIYEQIIEFCDAVEQEYKGLFADKTKWETTHLGYLHYYLIRYKITDERDFIIYHFRTSYRLFLEKFVMKRNFISN   53
ATCC6249           --MLQKFYDRAFVFLKLVEQEYASLGQSCSEWESLHLRFLLYYLIRFKIKSDRDFSLYHFRTAYRLYLDKLLQGETTLIQ   46
COL85/1862         --MLQKFYDREFVFLKLVEQEYASLGQSCSEWESLHLRFLLYYLIRFKIKSDRDFSLYHFRTAYRLYLDKLLQGGTTLIQ   44
bv.2str.SK95       --MLQKFYERAFVFLKLVEQEYASLCQSCAEWESLHLRFLLYYLIRFRIKSDKEFCLYHFQTAYRLYLDKFLLEGTSLN-   43
OP51               --MLQKFYDRAFVFLKLVEQEYASLGQSCSEWESLHLRFLLYYLIRFKIKSDRDFSLYHFKTAYRLYLDKFLQGGITLIQ   42
                      :*.:*::   *   :*:**     .  .**  *: :* ***.:: *   ::* :**::.:***:*:*::     :
```

Figure 29. Alignment of ComW homologs from *S. pneumoniae* and various strains of *S. mitis.* *S. pneumoniae* strain R6 ComW residues (accession number AAK98824.1) were aligned with corresponding residues of ComW sequences from *S. mitis* strains SK579 (EID31950.1), SK616 (EID23948.1), SK1073 (EGP70486.1), SK597 (EFO01026.1), SK1080 (EGP70101.1), SK1126 (KEQ34303.1), SK564 (EFN98169.1), 11/5 (EOB21306.1), bv.2 str.F0392 (EGR93223.1), ATCC6249 (EFM32340.1), COL85/1862 (KJQ61036.1), bv.2 str.SK95 (EGU63936.1), and OP51 (KJQ66299.1). Alignment was made using Clustal Omega with default parameters (Sievers et al. 2011). Green text, tracts of conserved residues; black text, regions of non-conserved residues. Species or strain name is on the left, sequence is in middle, percent identity to *S. pneumoniae* R6 sequence is on the right. Asterisks, identical residues; colons, conserved residues; periods, semi-conserved residues.

```
Strain/Species                                          Sequence                                          % Identity
                                                                                                            to R6
S.pneumo R6            MIMLQKIYEQMANFYDSIEEEYGPTFGDNFDWEHVHFKFLIYYLVRYGIGCRKDFIVYHYRVAYRLYLEKLVMNRGFISC     --
S.pneumo 430772        --MLQKIYEQMANFYDSIEEEYGPTFGDNFDWEHVHFKFLIYYLVRYGIGCRKDFIIYHYRVAYRLYLEKLVMNRGFISC     99
S.pneumo 2082239       --MLQKIYEQMANFYDSIEEEYGPTFGDNFDWEHIHFKFLIYYLVRYGIGCRRDFIVYHYRVAYRLYLEKLVMNRGFISC     97
S.pneumo SMRU1423      --MLQKIYEQMANFYDSIEEEYGPTFGDNFDWEYVHFKFLIYYLVRYGIGCHRDFIVYHYRVAYRLYLEKLVMNRGFISC     96
S.pneumo SMRU577       --MLKSIYEQMANFYDSIEEEYGPTFGDNFDWEHVHFKFLIYYLVRYGIGCRRDFIVYHYRVAYRLYLEKMIMKQGFISC     91
S.pneumo LMG1278       --MLQKIYEQMADFYRNIEEEYGTVFGDNFDWEHFHFKFLIYYLVRYGIGCRRDFIVYHYRVAYRLYLEKMIMKQGFISC     88
S.mitis  SK579         --MLQKIYEQMADFYRNIEEEYGTVFGDNFDWEHFHFKFLIYYLVRYGIGCRRDFIVYHYRVAYRLYLEKMIMNRGFISC     88
S.pneumo SMRU737       --MLQKIYEQMADFYRNIEEEYGTVFGDNFDWEHFHFKFLIYYLVRYGIGCRRDFIVYHYRVAYRLYLEKMIMNRGFISC     86
S.mitis  SK616         --MLQKIYEQMADFYRNIEEEYGTVFGDNFDWEHFHFKFLIYYLVRYGIGCRRDFIVYHYRVAYRLYLEKMIMKQGFISC     86
S.pseudopneumoniae     --MLQSIYEQMTDFYRNIEEEYGTVFGDNFDWEHFHFKFLIYYLVRYGIGCRRDFIVYHYRVAYRLYLEKMIMKQGFISC     86
S.mitis  SK1073        --MLQSIYDQMTDFYDSLEEEYATFFGDSWDWEHFHFKFLIYYLVRYGIGCRRDFIVYHYRVAYRLYLEKMIMNRGFISC     82
S.pneumo SMRU2014      --MLQSIYDQMTDFYDSIEEEYATFFGNSWDWEHFHFKFLIYYLVRYGIGCRRDFIVYHYRVAYRLYLEKLIMNQGFISC     81
S.pneumo 19F           --MLQNIYDQMTDFYDSIEEEYATFFDNSWNWEHFHFKFLIYYLVRYGIGCRRDFIVYHYRVAYRLYLEKLIMNQGFISC     79
S.mitis  SK597         --MLQNIYDQMTDFYDSIEEEYATFFGNSWDWEHFHFKFLIYYLVRYGIGCRRDFIVYHYRVTYRLYLEKMIMNRGFISC     79
S.mitis  SK1080        --MLQSIYDQMTDFYDSIEEEYAIFFGNSWDWEHFHFKFLIYYLVRYGIGCRRDFIVYHYRVAYRLYLEKMIMNRGFISC     78
S.pneumo SMRU946       --MLQNIYDQMTDFYDSIEEEYATFFDNSWEWEHFHFKFLIYYLVRYGIGCRRDFIVYHYRVAYRLYLEKMIMKQGFISC     77
S.pneumo SMRU824       --MLQSIYDQMTDFYDSIEEEYATFFDNSWEWEDFHFKFLIYYLVRYGIGCRRDFIVYHYRVAYRLYLEKLIMKQGFVAC     76
S.mitis  SK1126        --MLQSIYDQMTDFYDSIEEEYATFFGDSWDWEHFHFKFLIYYLFRYSIGNHRDFIVYHYRVAYRLYLEKMIMKQGFISC     76
S.mitis  SK564         MLTLQNIYDQMTDFYDSIEEEYATFFDNSWEWEHFHFKFLIYYLVRYGIGCRRDFIVYHYRVAYRLYLEKMIMKQGFISC     74
S.mitis  11/5          --MLQSIYNQMTDFYDSIEEEYATFFDNSWEWEHFHFKFLIYYLVQYDICRRRDFIVYHYRVAYRLYLEKMIMKQGFISC     72
S.mitis  bv.2str.F0392 --MFQDIYEQIIEFCDAVEQEYKGLFADKTKWETTHLGYLHYYLIRYKITDERDFIIYHFRVAYRLYLEKLIMKQGFVAC     53
S.anginosus            -MEIQEIYNQFRDYYGELEAEYAHCQKASMEWESLRLHYLIYYLIRYGIGEMKFFNAYHYRVAYRLYLEKMIMKQGFISC     49
S.mitis  ATCC6249      --MLQKFYDRAFVFLKLVEQEYASLGQSCSEWESLHLRFLLYYLIRFKIKSDRDFSLYHFRTSYRLFLEKFVMKRNFISN     46
S.oligofermentans      -MEIQEIYNQFRDYYGELEAEYAHCQKASMEWESLHLRYLIYYLMRYGIGEMKFFNAYHYRAAYRWYLQSLMLSSA----     46
S.cristatus            MMEIQEIYNQFRDYYGELEAEYAHCQKASTEWESLHLRYLIYYLMRYGIGEMKFFNAYHYRTAYRLYLDKLLQGETTLIQ     45
S.mitis  COL85/1862    --MLQKFYDREFVFLKLVEQEYASLGQSCSEWESLHLRFLLYYLIRFKIKSDRDFSLYHFRAAYRWYLQSLMLSST----     44
S.dentisani            --MLQKFYERAFVFLKLVEQEYVSLCQSSAEWESLHLRFLLYYLIRFKIKSDRDFSLYHFRAAYRWYLQSLMLSSA----     44
S.sinensis             -MKIQEIYLKYKGYYAEIEAEYSHCKKTSIEWETLHLRYLIYYLVRYNIGKMQFFNPYHYRTAYRLYLDKLLQGGTTLIQ     44
S.oralis               --MIEQIYEQYLDFYDVIEKEYSYLVDNDLEWEVFHLRFLLYYLVRYKLDIMHPLFSYHYRTAYRLYLEQLVAS------     43
S.infantis             --MIEQIYEQYLDFYDVIEKEYSYLVDNDLEWEVFHLRFLLYYLVRYKLDTMHPLFSYHYRACYRLYIEQLLISNDWVGG     43
S.mitis  bv.2str.SK95  --MLQKFYERAFVFLKLVEQEYASLCQSCAEWESLHLRFLLYYLIRFRIKSDKEFCLYHFRACYRLYIEQLLISNDWVGR     43
S.mitis  OP51          --MLQKFYDRAFVFLKLVEQEYASLGQSCSEWESLHLRFLLYYLIRFKIKSDRDFSLYHFQTAYRLYLDKFLLEGTSLN-     42
S.tigurinus            --MLQKFYDRAFVFLKLVEQEYASLGQSCAEWESLHLRFLLYYLIRFKIKSDRDFSLYHFKTAYRLYLDEFLQGGTTLIQ     40
                          ::.:* :   :   :* **        .**  :: :* ***.:: :   : :  **::. ** :::.::
```

Figure 30. Alignment of all ComW homologs from various species analyzed in this study. *S. pneumoniae* strain R6 ComW residues (accession number AAK98824.1) were aligned with corresponding residues of ComW sequences from *S. pneumoniae* 430772 (WP_05439300.1), *S. pneumoniae* 2082239 (EJG75956.1), *S. pneumoniae* SMRU1423 (COP90535.1), *S. pneumoniae* SMRU577 (COB21985.1), *S. pneumoniae* LMG1278 (CNA73777.1), *S. mitis* SK579 (EID31950.1), *S. pneumoniae* SMRU737 (CEY67237.1), *S. mitis* SK616 (EID23948.1), *S. pseudopneumoniae* (WP_049538978.1), *S. mitis* SK1073 (EGP70486.1), *S. pneumoniae* SMRU2014 (CJD59153.1), *S. pneumoniae* 19F (CEY35379.1), *S. mitis* SK597 (EFO01026.1), *S. mitis* SK1080 (EGP70101.1), *S. pneumoniae* SMRU946 (CIY80391.1), *S. pneumoniae* SMRU824 (CON90623.1), *S. mitis* SK1126 (KEQ34303.1), *S. mitis* SK564 (EFN98169.1), *S. mitis* bv.2 str.F0392 (EGR93223.1), *S. anginosus* (EJP26452.1), *S. mitis* ATCC6249 (EFM32340.1), *S. oligofermentans* (WP_015605924.1), *S. cristatus* (WP_005591530.1), *S. mitis* COL85/1862 (KJQ61036.1), *S. dentisani* (WP_038804352.1), *S. sinensis* (WP_037617413.1), *S. oralis* (WP_049503840.1), *S. infantis* (EFO53736.1), *S. mitis* bv.2 str.SK95 (EGU63936.1), *S. mitis* OP51 (KJQ66299.1), and *S. tigurinus* (EMG33941.1). Alignment was made using Clustal Omega using default parameters (Sievers et al. 2011). Purple text, regions of residue conservation, black text, regions of non-conservation; **bold**, possible motifs important to ComW function. Species and strain name are on the left, amino acid sequences are in the middle, percent identity to *S. pneumoniae* R6 is on the right. Asterisks, identical residues; colons, conserved residues; periods, semi-conserved residues; underline, highly conserved regions.

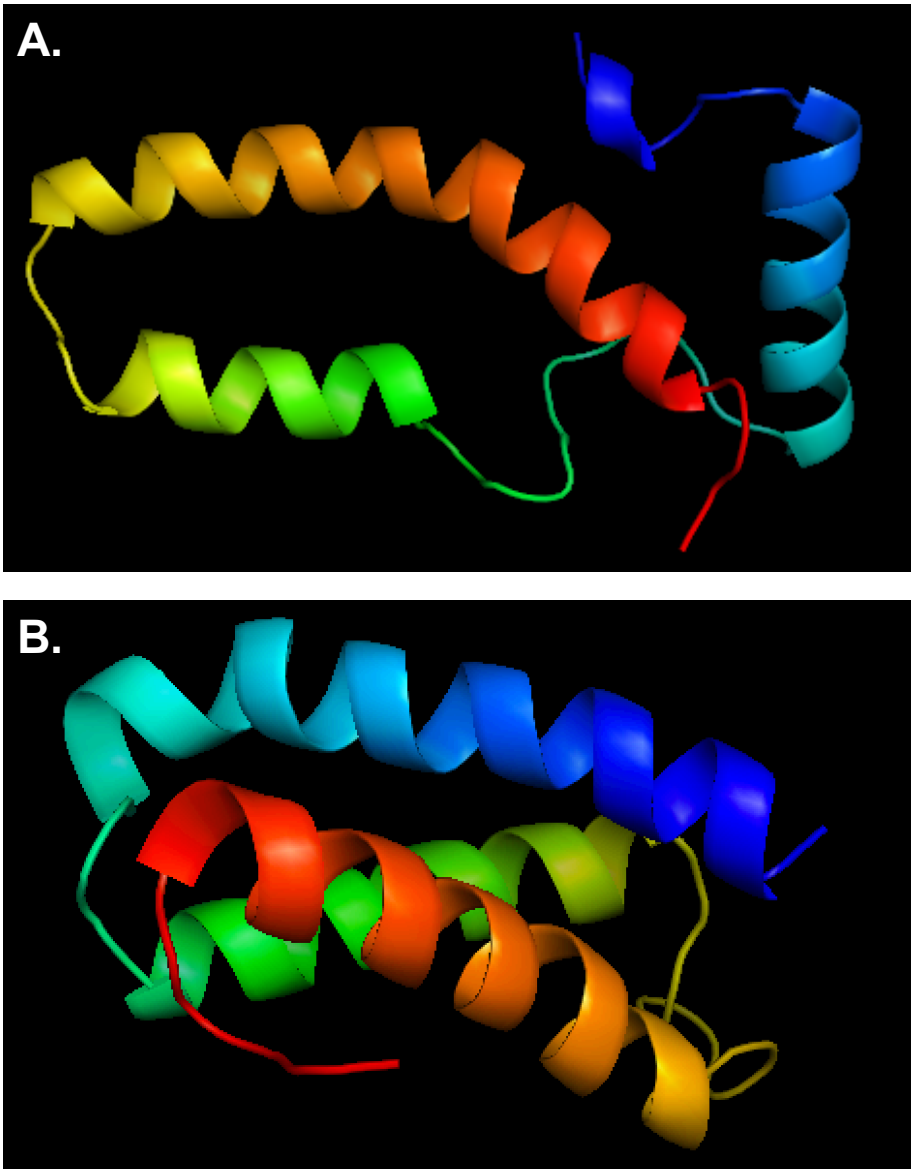


Figure 31. ComW ribbon structure prediction using Phyre2 and Robetta. Predicted 3D structures for ComW using the ComW sequence from *S. pneumoniae* (accession number AAK98824.1) using two different structural modeling prediction programs; Phyre2 (Kelley et al. 2015), and the Robetta Server (Kim et al. 2004, N. Inniss, personal communication). (A). Structure prediction using the Phyre2 program. This ribbon model is generated based on sequence similarity to other proteins with similar sequences and domains. ComW is predicted to have three helices with flexible regions in between the helices. (B). Structure prediction using the Robetta Server. This model also has three helices, but these helices are more tightly packed than those predicted by Phyre2. The range of colors from blue to red show the arrangement from the N- to C-terminus.

Robetta Server (Kim et al. 2004). The Robetta modeling system also predicts ComW to have three alpha helices that are more tightly packed than in the Phyre2 model prediction (N. Inniss, personal communication) (Fig. 31B). Aside from prediction modeling, crystallization and X-ray diffraction of ComW would provide a better understanding of the 3D structure of ComW. Furthermore, if a crystal structure is available, it could shed light on ComW function by revealing potential motifs or binding pockets.

The largest portion of this study explored the effects of single-base substitutions in the primary σ factor, σ^A, that partially restored transformation efficiency and late gene expression in *comW* mutants. Comparisons of the suppressor residues to previously studied residues in the primary σ factor implicated these residues in affinity for core RNAP (Nickels et al. 2005; Dove et al. 2003; Laurie et al. 2003). I predict that the bypass residues in σ^A* strains reduce σ^A affinity for core RNAP, increasing σ^X access for core without the need for ComW. However, we did not directly test whether these substitutions actually affected the affinity of σ^A for core RNAP. Measuring the differences in binding may reveal that these residues do, in fact, affect core binding. Similarly, it is important to know the binding affinity of σ^X compared to σ^A, and how this can improve in the presence of ComW if our hypothesis of ComW function is true. Increased σ^X binding in the presence of ComW would biochemically establish a function of ComW as a σ factor activator. Previous studies measuring alternative σ binding affinities to core RNAP using gel filtration or SPR (surface plasmon resonance) could serve as a model for such studies in *S. pneumoniae* (Maeda 2000; Kang et al. 1999).

ComW has at least two functions, to increase σ^X activity and σ^X stability (Sung & Morrison 2005). ComW is also suggested to improve σ^X affinity for core RNAP (Tovpeko & Morrison 2014). A new function proposed for ComW in this study is a potential role in the shut-off of late

gene expression. Every *comW* mutant strain, regardless of whether it had the WT or mutant σ^A, had persistent late gene expression. This may present an additional function for ComW, but it is possible that all the functions of ComW reflect a single role with different effects. For example, if ComW is an extremely labile σ factor activator that assists with loading of σ^X into core RNAP, this explains all the separate phenotypes observed. ComW binds to σ^X to load it onto core RNAP for a brief period during transformation, allowing for the transcription of late genes, as shown by its role in late gene transcription. Loading of σ^X onto core also allows for decreased σ^X proteolysis because it bound to core RNAP, as shown by ComW's role in increasing σ^X stability. Then, although it is unclear how ComW is involved, the correct shut-off mechanism is ineffective without ComW.

The overall roles and observed functions of ComW suggest that ComW acts as σ factor activator. While many alternative σ factors are negatively regulated by proteolysis and sequestration, the idea of a σ factor activator is not entirely novel (Österberg et al. 2011). The small protein Crl, discovered in *E. coli,* acts to promote σ^S binding to core RNAP (Gaal et al. 2006; Banta et al. 2013). However, Crl has only been identified in Gram-negative bacteria, even though it is found in many genera such as *Escherichia, Salmonella,* and *Pseudomonas* (Dong & Schellhorn 2010; Robbe-Saule et al. 2007). ComW is only found in the *Streptococci*, restricted to the mitis and and anginosus groups. If ComW is, in fact, a σ factor activator, its function would be novel in Gram-positive bacteria. The current studies about the function of ComW all point to its role as a protein responsible for promoting σ factor exchange during transformation (Fig. 32). Taken together, the date presented in this thesis suggests that ComW functions as a novel σ factor activator in *S. pneumoniae* during transformation.

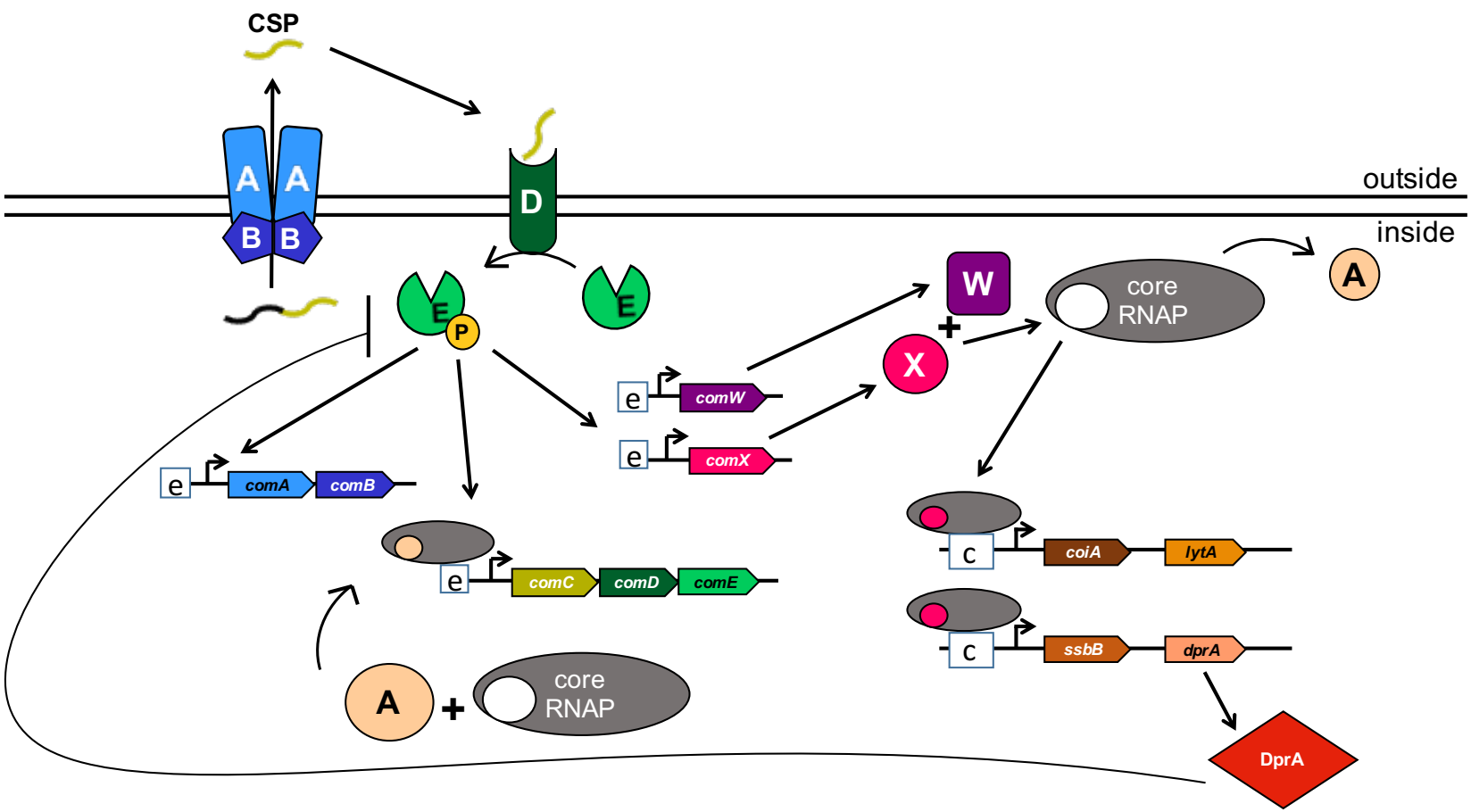


Figure 32. The transformation pathway in *S. pneumoniae* and proposed function for ComW. *comAB* and *comCDE* encode the machinery for exporting, cleaving, and sensing CSP. Upon sensing CSP, ComD phosphorylates ComE, which binds at the ComE-box to upregulate transcription of *comAB, comCDE*, *comX,* and *comW*. ComW promotes σ^X binding to core RNA polymerase. Then, σ^X can transcribe from the non-canonical promoter, or cin-box, upstream of late competence genes. DprA is a late protein, which negatively regulates ComE. Inside, inside of cell; outside, outside of cell; blue rectangles, ComA; small dark blue pentagons; ComB; green cylinder, ComD; light green circles, ComE; yellow circle, phosphate; gray oval, core RNAP, beige circle, σ^A; magenta circle, σ^X; purple square, ComW; brown, light brown, orange, peach, late genes; red, DprA; elongated pentagons, genes; bent arrow, promoter; curved or straight arrow, location of action; boxed e, ComE-box; boxed c, cinbox; double black line, cell wall/membrane.

CITED LITERATURE

Alloing, G. et al., 1998. Development of competence in Streptococcus pneumoniae: pheromone autoinduction and control of quorum sensing by the oligopeptide permease. *Molecular Microbiology*, 29(1), pp.75–83.

Attaiech, L. et al., 2011. Role of the single-stranded DNA-binding protein SsbB in pneumococcal transformation: maintenance of a reservoir for genetic plasticity. *PLoS genetics*, 7(6), p.e1002156.

Avery, O.T., 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. *Journal of Experimental Medicine*, 79(2), pp.137–158.

Bae, B. et al., 2013. Phage T7 Gp2 inhibition of Escherichia coli RNA polymerase involves misappropriation of σ70 domain 1.1. *Proceedings of the National Academy of Sciences of the United States of America*, 110(49), pp.19772–7.

Banta, A.B. et al., 2013. Key features of S required for specific recognition by Crl, a transcription factor promoting assembly of RNA polymerase holoenzyme. *Proceedings of the National Academy of Sciences*, 110(40), pp.15955–15960.

Barne, K.A. et al., 1997. Region 2.5 of the Escherichia coli RNA polymerase σ70 subunit is responsible for the recognition of the "extended −10" motif at promoters. *The EMBO Journal*, 16(13), pp.4034–4040.

Bartilson, M. et al., 2001. Differential fluorescence induction reveals Streptococcus pneumoniae loci regulated by competence stimulatory peptide. *Molecular Microbiology*, 39(1), pp.126–135.

Bassler, B.L. & Miller, M.B., 2013. Quorum Sensing. In *The Prokaryotes - Prokaryotic Communities and Ecophysiology*. Springer Berlin Heidelberg, pp. 495–509.

Bek-Thomsen, M. et al., 2008. Population diversity and dynamics of Streptococcus mitis, Streptococcus oralis, and Streptococcus infantis in the upper respiratory tracts of adults, determined by a nonculture strategy. *Infection and immunity*, 76(5), pp.1889–96.

Bergé, M.J. et al., 2013. Midcell recruitment of the DNA uptake and virulence nuclease, EndA, for pneumococcal transformation. *PLoS pathogens*, 9(9), p.e1003596.

Boutry, C. et al., 2012. Adaptor protein MecA is a negative regulator of the expression of late competence genes in Streptococcus thermophilus. *Journal of bacteriology*, 194(7), pp.1777–88.

Britton, R.A. et al., 2002. Genome-Wide Analysis of the Stationary-Phase Sigma Factor (Sigma-H) Regulon of Bacillus subtilis. *Journal of Bacteriology*, 184(17), pp.4881–4890.

Brueggemann, A.B. et al., 2007. Vaccine escape recombinants emerge after pneumococcal vaccination in the United States. *PLoS pathogens*, 3(11), p.e168.

Burgess, R.R. et al., 1969. Factor Stimulating Transcription by RNA Polymerase. *Nature*, 221(175), pp.43–46.

Campbell, E.A. et al., 2002. Structure of the Bacterial RNA Polymerase Promoter Specificity σ Subunit. *Molecular Cell*, 9(3), pp.527–539.

Campbell, E.A., Choi, S.Y. & Masure, H.R., 1998. A competence regulon in Streptococcus pneumoniae revealed by genomic analysis. *Molecular Microbiology*, 27(5), pp.929–939.

Cato, A. & Guild, W.R., 1968. Transformation and DNA size. *Journal of Molecular Biology*, 37(1), pp.157–178.

CDC, 2013. Antibiotic Resistance Threats in the United States. Available at: http://www.cdc.gov/drugresistance/threat-report-2013/ [Accessed January 28, 2016].

Chaguza, C., Cornick, J.E. & Everett, D.B., 2015. Mechanisms and impact of genetic recombination in the evolution of Streptococcus pneumoniae. *Computational and structural biotechnology journal*, 13, pp.241–7.

Chandler, M.S. & Morrison, D.A., 1988. Identification of two proteins encoded by com, a competence control locus of Streptococcus pneumoniae. *J. Bacteriol.*, 170(7), pp.3136–3141.

Charpentier, X. et al., 2011. Antibiotics and UV radiation induce competence for natural transformation in Legionella pneumophila. *Journal of bacteriology*, 193(5), pp.1114–21.

Chastanet, A. et al., 2001. Regulation of Streptococcus pneumoniae clp genes and their role in competence development and stress survival. *Journal of bacteriology*, 183(24), pp.7295–307.

Claverys, J.-P., Prudhomme, M. & Martin, B., 2006. Induction of Competence Regulons as a General Response to Stress in Gram-Positive Bacteria.

Claverys, J.P. & Lacks, S.A., 1986. Heteroduplex deoxyribonucleic acid base mismatch repair in bacteria. *Microbiological reviews*, 50(2), pp.133–65.

Contente, S. & Dubnau, D., 1979. Marker rescue transformation by linear plasmid DNA in Bacillus subtilis. *Plasmid*, 2(4), pp.555–571.

Dalia, A.B., McDonough, E. & Camilli, A., 2014. Multiplex genome editing by natural transformation. *Proceedings of the National Academy of Sciences of the United States of America*, 111(24), pp.8937–42.

Deretic, V. et al., 1994. Conversion of Pseudomonas aeruginosa to mucoidy in cystic fibrosis: environmental stress and regulation of bacterial virulence by alternative sigma factors. *Journal of bacteriology*, 176(10), pp.2773–80.

Dong, T. & Schellhorn, H.E., 2010. Role of RpoS in virulence of pathogens. *Infection and immunity*, 78(3), pp.887–97.

Dove, S.L., Darst, S.A. & Hochschild, A., 2003. Region 4 of σ as a target for transcription regulation. *Molecular Microbiology*, 48(4), pp.863–874.

Eldholm, V. et al., 2009. Fratricide in Streptococcus pneumoniae: contributions and role of the cell wall hydrolases CbpD, LytA and LytC. *Microbiology (Reading, England)*, 155(Pt 7), pp.2223–34.

Facklam, R., 2002. What Happened to the Streptococci: Overview of Taxonomic and Nomenclature Changes. *Clinical Microbiology Reviews*, 15(4), pp.613–630.

Farewell, A., Kvint, K. & Nyström, T., 1998. Negative regulation by RpoS: a case of sigma factor competition. *Molecular Microbiology*, 29(4), pp.1039–1051.

Finkel, S.E. & Kolter, R., 2001. DNA as a Nutrient: Novel Role for Bacterial Competence Gene Homologs. *Journal of Bacteriology*, 183(21), pp.6288–6293.

Fontaine, L. et al., 2010. A novel pheromone quorum-sensing system controls the development of natural competence in Streptococcus thermophilus and Streptococcus salivarius. *Journal of bacteriology*, 192(5), pp.1444–54.

Fontaine, L. et al., 2013. Mechanism of competence activation by the ComRS signalling system in streptococci. *Molecular microbiology*, 87(6), pp.1113–32.

Fontaine, L. et al., 2015. Regulation of competence for natural transformation in streptococci. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*, 33, pp.343–60.

Gaal, T. et al., 2006. Crl facilitates RNA polymerase holoenzyme formation. *Journal of bacteriology*, 188(22), pp.7966–70.

Ghei, O.K. & Lacks, S.A., 1967. Recovery of Donor Deoxyribonucleic Acid Marker Activity from Eclipse in Pneumococcal Transformation. *J. Bacteriol.*, 93(3), pp.816–829.

Golubchik, T. et al., 2012. Pneumococcal genome sequencing tracks a vaccine escape variant formed through a multi-fragment recombination event. *Nature genetics*, 44(3), pp.352–5.

Griffith, F., 1928. The Significance of Pneumococcal Types. *The Journal of hygiene*, 27(2), pp.113–59.

Grossman, A.D. et al., 1985. Mutations in the rpoH (htpR) gene of Escherichia coli K-12 phenotypically suppress a temperature-sensitive mutant defective in the sigma 70 subunit of RNA polymerase. *J. Bacteriol.*, 161(3), pp.939–943.

Gruber, T.M. & Gross, C.A., 2003. Multiple Sigma Subunits and the Partitioning of Bacterial Transcription Space.

Guerrini, F. & Fox, M.S., 1968. Genetic heterozygosity in pneumococcal transformation. *Proceedings of the National Academy of Sciences of the United States of America*, 59(2), pp.429–36.

Gurney, T. & Fox, M.S., 1968. Physical and genetic hybrids formed in bacterial transformation. *Journal of Molecular Biology*, 32(1), pp.83–100.

Haldenwang, W., 1995. The sigma factors of Bacillus subtilis. *Microbiol. Rev.*, 59(1), pp.1–30.

Hamilton, H.L. & Dillard, J.P., 2006. Natural transformation of Neisseria gonorrhoeae: from DNA donation to homologous recombination. *Molecular microbiology*, 59(2), pp.376–85.

Håvarstein, L.S. et al., 1996. Identification of the streptococcal competence-pheromone receptor. *Molecular Microbiology*, 21(4), pp.863–869.

Håvarstein, L.S., 2010. Increasing competence in the genus Streptococcus. *Molecular Microbiology*, 78(3), pp.541–544.

Havarstein, L.S., Coomaraswamy, G. & Morrison, D.A., 1995. An unmodified heptadecapeptide pheromone induces competence for genetic transformation in Streptococcus pneumoniae. *Proceedings of the National Academy of Sciences*, 92(24), pp.11140–11144.

Hernandez, J. V. & Cashel, M., 1995. Changes in Conserved Region 3 ofEscherichia coliσ70Mediate ppGpp-dependent FunctionsIn Vivo. *Journal of Molecular Biology*, 252(5), pp.536–549.

Hoskins, J. et al., 2001. Genome of the bacterium Streptococcus pneumoniae strain R6. *Journal of bacteriology*, 183(19), pp.5709–17.

Hui, F.M. & Morrison, D.A., 1991. Genetic transformation in Streptococcus pneumoniae: nucleotide sequence analysis shows comA, a gene required for competence induction, to be a member of the bacterial ATP-dependent transport protein family. *Journal of bacteriology*, 173(1), pp.372–81.

Hui, F.M., Zhou, L. & Morrison, D.A., 1995. Competence for genetic transformation in Streptococcus pneumoniae: organization of a regulatory locus with homology to two lactococcin A secretion genes. *Gene*, 153(1), pp.25–31.

Jarmer, H. et al., 2002. Transcriptome analysis documents induced competence of Bacillus subtilis during nitrogen limiting conditions. *FEMS Microbiology Letters*, 206(2), pp.197–200.

Jefferies, J.M.C. et al., 2004. Genetic analysis of diverse disease-causing pneumococci indicates high levels of diversity within serotypes and capsule switching. *Journal of clinical microbiology*, 42(12), pp.5681–8.

Johnsborg, O., Eldholm, V. & Håvarstein, L.S., 2007. Natural genetic transformation: prevalence, mechanisms and function. *Research in Microbiology*, 158(10), pp.767–778.

Johnston, C., Martin, B., et al., 2014. Bacterial transformation: distribution, shared mechanisms and divergent control. *Nature Reviews Microbiology*, 12, pp.181–196.

Johnston, C., Campo, N., et al., 2014. Streptococcus pneumoniae, le transformiste. *Trends in microbiology*, 22(3), pp.113–9.

Kang, J.G. et al., 1999. RsrA, an anti-sigma factor regulated by redox change. *The EMBO journal*, 18(15), pp.4292–8.

Kawamura, Y. et al., 1995. Determination of 16S rRNA sequences of Streptococcus mitis and Streptococcus gordonii and phylogenetic relationships among members of the genus Streptococcus. *International journal of systematic bacteriology*, 45(2), pp.406–8.

Kelley, L.A. et al., 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols*, 10(6), pp.845–858.

Kilian, M. et al., 2014. Parallel evolution of Streptococcus pneumoniae and Streptococcus mitis to pathogenic and mutualistic lifestyles. *mBio*, 5(4), pp.e01490–14.

Kim, D.E., Chivian, D. & Baker, D., 2004. Protein structure prediction and analysis using the Robetta server. *Nucleic acids research*, 32(Web Server issue), pp.W526–31.

Lacks, S.A., 1978. *Genetic Interaction and Gene Transfer: Report of a Symposium Held May 17-20, 1977, Issue 29* C. W. Anderson, ed., Brookhaven National Laboratory, Biology Department.

Lacks, S.A., 2004. Transformation. In T. Mitchell et al., eds. *The Pneumococcus*. Washington, DC: ASM Press, pp. 89–115.

Lancefield, R.C., 1933. A Serological differentiation of humand and other groups of Hemolytic Streptococci. *Journal of Experimental Medicine*, 57(4), pp.571–595.

Lancefield, R.C., 1962. Current knowledge of type-specific M antigens of group A streptococci. *Journal of immunology (Baltimore, Md. : 1950)*, 89(3), pp.307–13.

Lau, P.C.. et al., 2002. PCR ligation mutagenesis in transformable streptococci: application and efficiency. *Journal of Microbiological Methods*, 49(2), pp.193–205.

Laurie, A.D. et al., 2003. The role of the alarmone (p)ppGpp in sigma N competition for core RNA polymerase. *The Journal of biological chemistry*, 278(3), pp.1494–503.

Lee, M.S. & Morrison, D.A., 1999. Identification of a new regulator in Streptococcus pneumoniae linking quorum sensing to competence for genetic transformation. *Journal of bacteriology*, 181(16), pp.5004–16.

Lee, M.S., Seok, C. & Morrison, D.A., 1998. Insertion-Duplication Mutagenesis in Streptococcus pneumoniae: Targeting Fragment Length Is a Critical Parameter in Use as a Random Insertion Tool. *Appl. Envir. Microbiol.*, 64(12), pp.4796–4802.

Lee, Y.H. & Helmann, J.D., 2014. Mutations in the primary sigma factor σA and termination factor rho that reduce susceptibility to cell wall antibiotics. *Journal of bacteriology*, 196(21), pp.3700–11.

Li, Y.H. et al., 2001. Natural genetic transformation of Streptococcus mutans growing in biofilms. *Journal of bacteriology*, 183(3), pp.897–908.

Lonetto, M., Gribskov, M. & Gross, C.A., 1992. The sigma 70 family: sequence conservation and evolutionary relationships. *Journal of bacteriology*, 174(12), pp.3843–9.

Lonetto, M.A. et al., 1998. Identification of a contact site for different transcription activators in region 4 of the Escherichia coli RNA polymerase sigma70 subunit. *Journal of molecular biology*, 284(5), pp.1353–65.

Lorenz, M.G. & Wackernagel, W., 1994. Bacterial gene transfer by natural genetic transformation in the environment. *Microbiological reviews*, 58(3), pp.563–602.

Louarn, J.-M. & Sicard, A.M., 1968. Transmission of genetic information during transformation in. *Biochemical and Biophysical Research Communications*, 30(6), pp.683–689.

Luo, P., Li, H. & Morrison, D.A., 2003. ComX is a unique link between multiple quorum sensing outputs and competence in Streptococcus pneumoniae. *Molecular Microbiology*, 50(2), pp.623–633.

Luo, P., Li, H. & Morrison, D.A., 2004. Identification of ComW as a new component in the regulation of genetic transformation in Streptococcus pneumoniae. *Molecular microbiology*, 54(1), pp.172–83.

Luo, P. & Morrison, D.A., 2003. Transient association of an alternative sigma factor, ComX, with RNA polymerase during the period of competence for genetic transformation in Streptococcus pneumoniae. *Journal of bacteriology*, 185(1), pp.349–58.

Maeda, H., 2000. Competition among seven Escherichia coli sigma subunits: relative binding affinities to the core RNA polymerase. *Nucleic Acids Research*, 28(18), pp.3497–3503.

Marks, L.R. et al., 2014. Streptococcus pyogenes biofilm growth in vitro and in vivo and its role in colonization, virulence, and genetic exchange. *The Journal of infectious diseases*, 210(1), pp.25–34.

Martin, B. et al., 2013. ComE/ComE~P interplay dictates activation or extinction status of pneumococcal X-state (competence). *Molecular microbiology*, 87(2), pp.394–411.

Martin, B. et al., 2010. Expression and maintenance of ComD-ComE, the two-component signal-transduction system that controls competence of Streptococcus pneumoniae. *Molecular microbiology*, 75(6), pp.1513–28.

Martin, B. et al., 2006. Independent evolution of competence regulatory cascades in streptococci? *Trends in microbiology*, 14(8), pp.339–45.

Martín, V. et al., 2015. Streptococcal Diversity of Human Milk and Comparison of Different Methods for the Taxonomic Identification of Streptococci. *Journal of human lactation : official journal of International Lactation Consultant Association*, p.0890334415597901–.

Mashburn-Warren, L., Morrison, D.A. & Federle, M.J., 2010. A novel double-tryptophan peptide pheromone controls competence in Streptococcus spp. via an Rgg regulator. *Molecular microbiology*, 78(3), pp.589–606.

Mejean, V. & Claverys, J.P., 1993. DNA processing during entry in transformation of Streptococcus pneumoniae. *J. Biol. Chem.*, 268(8), pp.5594–5599.

Miller, M.B. & Bassler, B.L., 2003. Quorum Sensing in Bacteria.

Mirouze, N. et al., 2013. Direct involvement of DprA, the transformation-dedicated RecA loader, in the shut-off of pneumococcal competence. *Proceedings of the National Academy of Sciences of the United States of America*, 110(11), pp.E1035–44.

Miyazaki, E. et al., 1999. The Staphylococcus aureus rsbW (orf159) Gene Encodes an Anti-Sigma Factor of SigB. *J. Bacteriol.*, 181(9), pp.2846–2851.

Morrison, D.A., 2007. Genetic transformation in Streptococcus pneumoniae: strategies for genome modification. In R. Hackenbeck & C. Singh, eds. *Molecular Biology of Streptococci*. Norfolk: Horizon Scientific Press, pp. 3–24.

Morrison, D.A. et al., 2015. Genome editing by natural genetic transformation in Streptococcus mutans. *Journal of microbiological methods*, 119, pp.134–41.

Morrison, D.A. et al., 2007. Identification of the major protein component of the pneumococcal eclipse complex. *Journal of bacteriology*, 189(17), pp.6497–500.

Morrison, D.A., Guédon, E. & Renault, P., 2013. Competence for natural genetic transformation in the Streptococcus bovis group streptococci S. infantarius and S. macedonicus. *Journal of bacteriology*, 195(11), pp.2612–20.

Morrison, D.A. & Guild, W.R., 1972. Activity of Deoxyribonucleic Acid Fragments of Defined Size in Bacillus subtilis Transformation. *J. Bacteriol.*, 112(1), pp.220–223.

Morrison, D.A. & Jaurin, B., 1990. Streptococcus pneumoniae possesses canonical Escherichia coli (sigma 70) promoters. *Molecular Microbiology*, 4(7), pp.1143–1152.

Morrison, D.A. & Mannarelli, B., 1979. Transformation in pneumococcus: nuclease resistance of deoxyribonucleic acid in the eclipse complex. *Journal of bacteriology*, 140(2), pp.655–65.

Mortier-Barriere, I. et al., 1997. Control of Recombination Rate During Transformation of Streptococcus pneumoniae : An Overview. *Microbial Drug Resistance*, 3(3), pp.233–242.

Muñoz, R., Bustamante, M. & de la Campa, A.G., 1995. Ser-127-to-Leu substitution in the DNA gyrase B subunit of Streptococcus pneumoniae is implicated in novobiocin resistance. *Journal of bacteriology*, 177(14), pp.4166–70.

Murakami, K.S., Masuda, S., Campbell, E.A., et al., 2002. Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science (New York, N.Y.)*, 296(5571), pp.1285–90.

Murakami, K.S. & Darst, S.A., 2003. Bacterial RNA polymerases: the wholo story. *Current Opinion in Structural Biology*, 13(1), pp.31–39.

Murakami, K.S., Masuda, S. & Darst, S.A., 2002. Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 A resolution. *Science (New York, N.Y.)*, 296(5571), pp.1280–4.

Nickels, B.E. et al., 2005. The interaction between sigma70 and the beta-flap of Escherichia coli RNA polymerase inhibits extension of nascent RNA during early elongation. *Proceedings of the National Academy of Sciences of the United States of America*, 102(12), pp.4488–93.

Ochman, H., Lawrence, J.G. & Groisman, E.A., 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature*, 405(6784), pp.299–304.

Oggioni, M. & Morrison, D., 2008. Cooperative regulation of competence development in Streptococcus pneumoniae: cell-to-cell signaling via a peptide pheromone and an alternative sigma factor. In W. SC & B. Bassler, eds. *Chemical communication among bacteria*. Washington, DC: ASM Press, pp. 345–362.

Österberg, S., del Peso-Santos, T. & Shingler, V., 2011. Regulation of alternative sigma factor use. *Annual review of microbiology*, 65, pp.37–55.

Paget, M.S.B. & Helmann, J.D., 2003. The sigma70 family of sigma factors. *Genome biology*, 4(1), p.203.

Pestova, E. V., Håvarstein, L.S. & Morrison, D.A., 1996. Regulation of competence for genetic transformation in Streptococcus pneumoniae by an auto-induced peptide pheromone and a two-component regulatory system. *Molecular Microbiology*, 21(4), pp.853–862.

Peterson, S. et al., 2000. Gene Expression Analysis of the Streptococcus pneumoniae Competence Regulons by Use of DNA Microarrays. *Journal of Bacteriology*, 182(21), pp.6192–6202.

Peterson, S.N. et al., 2004. Identification of competence pheromone responsive genes in Streptococcus pneumoniae by use of DNA microarrays. *Molecular Microbiology*, 51(4), pp.1051–1070.

Piotrowski, A., Luo, P. & Morrison, D.A., 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. *Journal of bacteriology*, 191(10), pp.3359–66.

Polz, M.F., Alm, E.J. & Hanage, W.P., 2013. Horizontal gene transfer and the evolution of bacterial and archaeal population structure. *Trends in genetics : TIG*, 29(3), pp.170–5.

Porter, R.D. & Guild, W.R., 1969. Number of Transformable Units Per Cell in Diplococcus pneumoniae. *J. Bacteriol.*, 97(3), pp.1033–1035.

Pozzi, G. et al., 1996. Competence for genetic transformation in encapsulated strains of Streptococcus pneumoniae: two allelic variants of the peptide pheromone. *Journal of bacteriology*, 178(20), pp.6087–90.

Puyet, A., Greenberg, B. & Lacks, S.A., 1990. Genetic and structural characterization of endA. A membrane-bound nuclease required for transformation of Streptococcus pneumoniae. *Journal of molecular biology*, 213(4), pp.727–38.

Reck, M., Tomasch, J. & Wagner-Döbler, I., 2015. The Alternative Sigma Factor SigX Controls Bacteriocin Synthesis and Competence, the Two Quorum Sensing Regulated Traits in Streptococcus mutans. *PLoS genetics*, 11(7), p.e1005353.

Redfield, R.J., 1993. Genes for Breakfast: The Have-Your-Cake and-Eat-lt-Too of Bacterial Transformation. *J. Hered.*, 84(5), pp.400–404.

Reznikoff, W.S. et al., 1985. The regulation of transcription initiation in bacteria. *Annual review of genetics*, 19, pp.355–87.

Richards, V.P. et al., 2014. Phylogenomics and the dynamic genome evolution of the genus Streptococcus. *Genome biology and evolution*, 6(4), pp.741–53.

Robbe-Saule, V. et al., 2007. Physiological effects of Crl in Salmonella are modulated by sigmaS level and promoter specificity. *Journal of bacteriology*, 189(8), pp.2976–87.

Robertson, G.T. et al., 2002. Global Transcriptional Analysis of clpP Mutations of Type 2 Streptococcus pneumoniae and Their Effects on Physiology and Virulence. *Journal of Bacteriology*, 184(13), pp.3508–3520.

Rodriguez, A.M. et al., 2011. Physiological and molecular characterization of genetic competence in Streptococcus sanguinis. *Molecular oral microbiology*, 26(2), pp.99–116.

Schleifer, K.H. & Kilpper-Balz, R., 1984. Transfer of Streptococcus faecalis and Streptococcus faecium to the Genus Enterococcus nom. rev. as Enterococcus faecalis comb. nov. and Enterococcus faecium comb. nov. *International Journal of Systematic Bacteriology*, 34(1), pp.31–34.

Schöbel, S. et al., 2004. The Bacillus subtilis sigmaW anti-sigma factor RsiW is degraded by intramembrane proteolysis through YluC. *Molecular microbiology*, 52(4), pp.1091–105.

Severinov, K. et al., 1994. The sigma subunit conserved region 3 is part of "5'-face" of active center of Escherichia coli RNA polymerase. *J. Biol. Chem.*, 269(33), pp.20826–20828.

Sievers, F. et al., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*, 7(1), p.539.

Straus, D., Walter, W. & Gross, C.A., 1990. DnaK, DnaJ, and GrpE heat shock proteins negatively regulate heat shock gene expression by controlling the synthesis and stability of sigma 32. *Genes & Development*, 4(12a), pp.2202–2209.

Sung, C.K. et al., 2001. An rpsL cassette, janus, for gene replacement through negative selection in Streptococcus pneumoniae. *Applied and environmental microbiology*, 67(11), pp.5190–6.

Sung, C.K. & Morrison, D.A., 2005. Two distinct functions of ComW in stabilization and activation of the alternative sigma factor ComX in Streptococcus pneumoniae. *Journal of bacteriology*, 187(9), pp.3052–61.

Swartley, J.S. et al., 1997. Capsule switching of Neisseria meningitidis. *Proceedings of the National Academy of Sciences*, 94(1), pp.271–276.

Tao, L. et al., 1993. Transport of Sugars, Including Sucrose, by the msm Transport System of Streptococcus mutans. *Journal of Dental Research*, 72(10), pp.1386–1390.

Tiraby, G., Fox, M.S. & Bernheimer, H., 1975. Marker discrimination in deoxyribonucleic acid-mediated transformation of various Pneumococcus strains. *J. Bacteriol.*, 121(2), pp.608–618.

Tiraby, J.-G. & Fox, M.S., 1973. Marker Discrimination in Transformation and Mutation of Pneumococcus. *Proceedings of the National Academy of Sciences*, 70(12), pp.3541–3545.

Tomasz, A. & Hotchkiss, R.D., 1964. REGULATION OF THE TRANSFORMABILITY OF PHEUMOCOCCAL CULTURES BY MACROMOLECULAR CELL PRODUCTS. *Proceedings of the National Academy of Sciences of the United States of America*, 51, pp.480–7.

Tovpeko, Y. & Morrison, D.A., 2014. Competence for genetic transformation in Streptococcus pneumoniae: mutations in σA bypass the comW requirement. *Journal of bacteriology*, 196(21), pp.3724–34.

Tsakalidou, E. et al., Identification of streptococci from Greek Kasseri cheese and description of Streptococcus macedonicus sp. nov. *International journal of systematic bacteriology*, 48(2), pp.519–527.

Tu, X. et al., 2006. The PhoP/PhoQ two-component system stabilizes the alternative sigma factor RpoS in Salmonella enterica. *Proceedings of the National Academy of Sciences of the United States of America*, 103(36), pp.13503–8.

Vassylyev, D.G. et al., 2002. Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution. *Nature*, 417(6890), pp.712–9.

Vickerman, M.M. et al., 2007. Genome-wide transcriptional changes in Streptococcus gordonii in response to competence signaling peptide. *Journal of bacteriology*, 189(21), pp.7799–807.

Voelker, U., Dufour, A. & Haldenwang, W.G., 1995. The Bacillus subtilis rsbU gene product is necessary for RsbX-dependent regulation of sigma B. *Journal of bacteriology*, 177(1), pp.114–22.

Ween, O., Gaustad, P. & Havarstein, L.S., 1999. Identification of DNA binding sites for ComE, a key regulator of natural competence in Streptococcus pneumoniae. *Molecular Microbiology*, 33(4), pp.817–827.

Weng, L., Biswas, I. & Morrison, D.A., 2009. A self-deleting Cre-lox-ermAM cassette, Cheshire, for marker-less gene deletion in Streptococcus pneumoniae. *Journal of microbiological methods*, 79(3), pp.353–7.

Weng, L., Piotrowski, A. & Morrison, D.A., 2013. Exit from competence for genetic transformation in Streptococcus pneumoniae is regulated at multiple levels. *PloS one*, 8(5), p.e64197.

Wilson, C. & Dombroski, A.J., 1997. Region 1 of σ70 is required for efficient isomerization and initiation of transcription by Escherichia coli RNA polymerase. *Journal of Molecular Biology*, 267(1), pp.60–74.

Woodbury, R.L., Wang, X. & Moran, C.P., 2006. Sigma X induces competence gene expression in Streptococcus pyogenes. *Research in microbiology*, 157(9), pp.851–6.

Wyres, K.L. et al., 2013. Pneumococcal capsular switching: a historical perspective. *The Journal of infectious diseases*, 207(3), pp.439–49.

Yin, Y. et al., 2013. Evidence for sigma factor competition in the regulation of alginate production by Pseudomonas aeruginosa. *PloS one*, 8(8), p.e72329.

Yother, J., McDaniel, L.S. & Briles, D.E., 1986. Transformation of encapsulated Streptococcus pneumoniae. *J. Bacteriol.*, 168(3), pp.1463–1465.

Zhou, Y. & Gottesman, S., 1998. Regulation of proteolysis of the stationary-phase sigma factor RpoS. *Journal of bacteriology*, 180(5), pp.1154–8.

APPENDICES

APPENDIX A

Table XI. Non-synonymous base changes in sequenced bypass mutants.

R6 Position[a]	R6 Base	SNP[b]	SNP Reads[c]	Freq.[d]	Gene[e]	Coding region change[f]	Amino acid change[g]	Strain[h]	Unique Gene
79856	T	C	194	89	spr0073, spr0074	NP_357667.1:c.[185A>G]; NP_357668.1:c.[197T>C]	NP_357667.1:p.[His62Arg]; NP_357668.1:p.[Val66Ala]	SCT-4	1
137972	A	G	182	99	spr0129	NP_357723.1:c.88A>G	NP_357723.1:p.Asn30Asp	ALT-4	2
172808	T	A	99	95	ribD	NP_357758.1:c.952A>T	NP_357758.1:p.Ile318Phe	TXT-4	3
188052	T	C	124	98	cls	NP_357774.1:c.323T>C	NP_357774.1:p.Phe108Ser	DET-5	4
204465	A	G	186	99	rpsE	NP_357800.1:c.145A>G	NP_357800.1:p.Thr49Ala	MAT-5	5
206465	T	C	126	97	secY	NP_357803.1:c.857T>C	NP_357803.1:p.Ile286Thr	GAT-1	6
237689	G	A	218	100	proS	NP_357837.1:c.916G>A	NP_357837.1:p.Ala306Thr	MTT-5	7
239704	G	A	218	100	bglA	NP_357838.1:c.978G>A	NP_357838.1:p.Met326Ile	FLT-4	8
298527	A	G	110	99	spr0297	NP_357891.1:c.1576A>G	NP_357891.1:p.Ile526Val	DET-5	9
341113	G	A	118	91	gnd	NP_357929.1:c.1381G>A	NP_357929.1:p.Asp461Asn	GAT-1	10
343401	T	C	193	98	mvk	NP_357932.1:c.308T>C	NP_357932.1:p.Val103Ala	NVT-1	11
346404	G	A	58	94	fni	NP_357935.1:c.520G>A	NP_357935.1:p.Val174Ile	NYT-2	12
362984	C	T	58	95	recD	NP_357957.1:c.658G>A	NP_357957.1:p.Ala220Thr	NYT-2	13
368383	A	G	143	100	mutS2	NP_357962.1:c.1973A>G	NP_357962.1:p.Gln658Arg	MAT-5	14
395308	C	G	171	98	asp23	NP_357993.1:c.188C>G	NP_357993.1:p.Ala63Gly	FLT-4	15
429371	G	A	136	99	spr0429	NP_358023.1:c.49G>A	NP_358023.1:p.Ala17Thr	MAT-5	16
435792	G	A	182	95	spr0436	NP_358030.1:c.324G>A	NP_358030.1:p.Trp108*	NVT-1	17
498602	G	A	91	98	spr0499	NP_358093.1:c.787G>A	NP_358093.1:p.Ala263Thr	GAT-1	18

a. Nucleotide position in R6 genome

b. Single nucleotide polymorphism (SNP) found in bypass mutant

c. Number of sequence reads containing SNP

d. % of reads with SNP base

e. Gene in R6 containing SNP

f. Accession number: base change

g. Accession number: amino acid change

h. Suppressor mutant strain in which the SNP is found

APPENDIX A (continued)

Table XI. (continued)

R6 Position[a]	R6 Base	SNP[b]	SNP Reads[c]	Freq.[d]	Gene[e]	Coding region change[f]	Amino acid change[g]	Strain[h]	Unique Gene
517931	G	A	51	91	pnpA	NP_358110.1:c.2236G>A	NP_358110.1:p.Gly746Ser	NYT-1	19
519232	G	A	131	98	spr0518	NP_358112.1:c.565G>A	NP_358112.1:p.Ala189Thr	NYT-1	20
592126	A	G	90	99	zmpB	NP_358175.1:c.680A>G	NP_358175.1:p.Asp227Gly	OHT-1	21
606409	T	C	105	100	spr0592	NP_358186.1:c.329T>C	NP_358186.1:p.Leu110Pro	MTT-5	22
649209	C	T	86	97	ctpA	NP_358235.1:c.2224C>T	NP_358235.1:p.Leu742Phe	GAT-1	23
715042	G	T	78	99	thiJ	NP_358307.1:c.442G>T	NP_358307.1:p.Gly148Cys	GAT-1	24
720738	G	A	74	100	transposase_D	NP_358312.1:c.131G>A	NP_358312.1:p.Arg44His	GAT-1	25
826714	A	G	104	99	spr0824	NP_358418.1:c.562T>C	NP_358418.1:p.Tyr188His	ILT-1	26
965792	C	T	112	100	rpoD	NP_358573.1:c.512C>T	NP_358573.1:p.Ala171Val	DET-5	27
965792	C	T	119	98	rpoD	NP_358573.1:c.512C>T	NP_358573.1:p.Ala171Val	MAT-5	27
965792	C	T	152	99	rpoD	NP_358573.1:c.512C>T	NP_358573.1:p.Ala171Val	ALT-4	27
965792	C	T	186	99	rpoD	NP_358573.1:c.512C>T	NP_358573.1:p.Ala171Val	NCT-2	27
966227	G	A	59	100	rpoD	NP_358573.1:c.947G>A	NP_358573.1:p.Arg316His	NYT-2	27
966227	G	A	70	100	rpoD	NP_358573.1:c.947G>A	NP_358573.1:p.Arg316His	TXT-4	27
966227	G	A	96	100	rpoD	NP_358573.1:c.947G>A	NP_358573.1:p.Arg316His	ILT-1	27
966227	G	A	126	100	rpoD	NP_358573.1:c.947G>A	NP_358573.1:p.Arg316His	GAT-1	27
966227	G	A	152	100	rpoD	NP_358573.1:c.947G>A	NP_358573.1:p.Arg316His	MTT-5	27
966227	G	A	228	100	rpoD	NP_358573.1:c.947G>A	NP_358573.1:p.Arg316His	NVT-1	27
966227	G	A	248	100	rpoD	NP_358573.1:c.947G>A	NP_358573.1:p.Arg316His	SCT-4	27
966344	G	A	140	100	rpoD	NP_358573.1:c.1064G>A	NP_358573.1:p.Arg355His	OHT-1	27

a. Nucleotide position in R6 genome

b. Single nucleotide polymorphism (SNP) found in bypass mutant

c. Number of sequence reads containing SNP

d. % of reads with SNP base

e. Gene in R6 containing SNP

f. Accession number: base change

g. Accession number: amino acid change

h. Suppressor mutant strain in which the SNP is found

APPENDIX A (continued)

Table XI. (continued)

R6 Position[a]	R6 Base	SNP[b]	SNP Reads[c]	Freq.[d]	Gene[e]	Coding region change[f]	Amino acid change[g]	Strain[h]	Unique Gene
966344	G	A	141	100	rpoD	NP_358573.1:c.1064G>A	NP_358573.1:p.Arg355His	FLT-4	27
966367	C	T	87	99	rpoD	NP_358573.1:c.1087C>T	NP_358573.1:p.Leu363Phe	NYT-1	27
1072831	T	C	190	99	lacC	NP_358667.1:c.847A>G	NP_358667.1:p.Asn283Asp	FLT-4	28
1091608	G	A	161	99	spr1094	NP_358687.1:c.569C>T	NP_358687.1:p.Ala190Val	NCT-2	29
1102165	T	C	165	93	spnII	NP_358694.1:c.587A>G	NP_358694.1:p.Asp196Gly	NCT-2	30
1119841	G	A	91	99	glnP	NP_358713.1:c.184G>A	NP_358713.1:p.Gly62Ser	NYT-1	31
1128529	T	C	242	100	smc	NP_358719.1:c.2146A>G	NP_358719.1:p.Lys716Glu	NCT-2	32
1128802	C	T	45	98	smc	NP_358719.1:c.1873G>A	NP_358719.1:p.Ala625Thr	NYT-2	32
1143200	C	T	119	94	topA	NP_358734.1:c.613G>A	NP_358734.1:p.Glu205Lys	NYT-1	33
1212019	G	A	193	98	ABC-N/P	NP_358809.1:c.1535C>T	NP_358809.1:p.Ala512Val	NCT-2	34
1233121	T	C	104	99	spr1236	NP_358829.1:c.1003A>G	NP_358829.1:p.Asn335Asp	GAT-1	35
1253298	C	T	145	99	pstB	NP_358847.1:c.529G>A	NP_358847.1:p.Ala177Thr	ALT-4	36
1299392	C	A	41	100	spr1307	NP_358900.1:c.130G>T	NP_358900.1:p.Ala44Ser	NYT-2	37
1375972	C	T	236	99	asnC	NP_358990.1:c.1138G>A	NP_358990.1:p.Asp380Asn	SCT-4	38
1401571	C	T	109	98	dapB	NP_359007.1:c.334G>A	NP_359007.1:p.Gly112Ser	GAT-1	39
1464129	C	T	107	95	relA	NP_359080.1:c.1029G>A	NP_359080.1:p.Trp343*	OHT-1	40
1467154	G	A	119	100	pepO	NP_359084.1:c.1643C>T	NP_359084.1:p.Ala548Val	DET-5	41
1483714	T	C	114	98	divIVA	NP_359098.1:c.692A>G	NP_359098.1:p.Glu231Gly	DET-5	42
1519669	A	G	49	98	axe1	NP_359131.1:c.728T>C	NP_359131.1:p.Ile243Thr	NYT-2	43
1520053	C	T	115	97	axe1	NP_359131.1:c.344G>A	NP_359131.1:p.Arg115Gln	MTT-5	43

a. Nucleotide position in R6 genome

b. Single nucleotide polymorphism (SNP) found in bypass mutant

c. Number of sequence reads containing SNP

d. % of reads with SNP base

e. Gene in R6 containing SNP

f. Accession number: base change

g. Accession number: amino acid change

h. Suppressor mutant strain in which the SNP is found

APPENDIX A (continued)

Table XI. (continued)

R6 Position[a]	R6 Base	SNP[b]	SNP Reads[c]	Freq.[d]	Gene[e]	Coding region change[f]	Amino acid change[g]	Strain[h]	Unique Gene
1541906	G	A	175	99	natB	NP_359154.1:c.427C>T	NP_359154.1:p.Gln143*	OHT-1	44
1542870	A	G	127	100	natA	NP_359155.1:c.353T>C	NP_359155.1:p.Val118Ala	NYT-1	45
1558537	C	T	124	98	pppL	NP_359170.1:c.284G>A	NP_359170.1:p.Arg95Lys	MAT-5	46
1616338	G	A	165	99	ABC-NBD	NP_359236.1:c.650C>T	NP_359236.1:p.Pro217Leu	DET-5	47
1688914	C	T	182	91	msmR	NP_359306.1:c.776C>T	NP_359306.1:p.Ala259Val	NVT-1	50
1754472	C	T	166	100	endA	NP_359371.1:c.229G>A	NP_359371.1:p.Ala77Thr	FLT-4	51
1777381	A	G	119	99	spr1806	NP_359398.1:c.425T>C	NP_359398.1:p.Val142Ala	NYT-1	52
1883328	T	C	148	100	hipO	NP_359497.1:c.737A>G	NP_359497.1:p.His246Arg	GAT-1	53

a. Nucleotide position in R6 genome

b. Single nucleotide polymorphism (SNP) found in bypass mutant

c. Number of sequence reads containing SNP

d. % of reads with SNP base

e. Gene in R6 containing SNP

f. Accession number: base change

g. Accession number: amino acid change

h. Suppressor mutant strain in which the SNP is found

VITA

YANINA TOVPEKO, B.S.
yanina.tovpeko@gmail.com

EDUCATION

University of Illinois at Chicago, Chicago, IL **May 2016**
Doctor of Philosophy in Molecular, Cell, and Developmental Biology
Area of Specialization: Microbiology and Microbial Genetics

Syracuse University, Syracuse, NY **2011**
Bachelor of Science in Biology, Minor in Psychology
Distinction In Biology, Renée Crown University Honors Program, S*umma Cum Laude*, Phi Beta Kappa

AWARDS AND GRANTS

Best Graduate Student Presentation - 59th Wind River Conference on Prokaryotic Biology **2015**
University Research Achievement Award - University of Illinois at Chicago, Biological Sciences **2015**
University Excellence in Teaching Award - University of Illinois at Chicago, Microbiology Laboratory **2015**
Excellence in Undergraduate Mentoring - Honorable Mention, University of Illinois at Chicago **2015**
Bodmer International Travel Grant - University of Illinois at Chicago **2013**
LAS PhD Student Travel Award - University of Illinois at Chicago **2013**
Graduate College Student Presenter Award – University of Illinois at Chicago **2013, 2014**
Biological Sciences Departmental Travel Award - University of Illinois at Chicago **2012, 2013, 2014, 2015**

RESEARCH EXPERIENCE

Doctoral Research; Advisor: Donald A. Morrison, Ph.D. **August 2011 – May 2016**
University of Illinois at Chicago, Biological Sciences, Molecular, Cell, and Developmental Biology
Dissertation: Mutations in the primary sigma factor bypass the *comW* requirement for genetic transformation in *Streptococcus pneumoniae*.

- Elucidated the role of a small protein of unknown function that is required for natural transformation in *S. pneumoniae* and identified its role in interactions with the primary σ factor.

Undergraduate Researcher; Advisor: Kari Segraves, Ph.D. **August 2010 – May 2011**
Syracuse University, Department of Biology
Title: Mutualisms of *Yucca filamentosa* and the Yucca Moth

- Studied plant-moth interactions in *Yucca filamentosa*, microsatellite analyses of plant paternity

Undergraduate Honors Thesis Research; Advisor: Robert R. Lebel, MD **August 2009 – May 2011**
SUNY Upstate Medical University, Center for Behavioral Genetics
Thesis: The Basis and Implications of Genetic Counseling in Patient Testing Choice

- Conducted independent research concerning pre-natal testing choice in high-risk pregnancies

PEER-REVIEWED PUBLICATIONS

Tovpeko, Y. and Morrison, D.A. Primary σ factor mutations bypass the ComW requirement for late gene expression during genetic transformation in *Streptococcus pneumoniae*. (manuscript in preparation)
Junges, R., Khan, R., **Tovpeko, Y.**, Åmdal, H., Petersen, F.C., and Morrison, D.A. Genome editing in competent *streptococci*. Methods in Microbiology – Oral Biology. 2nd Edition, New Zealand Press. 2. (2015).

Tovpeko, Y. and Morrison, D.A. Competence for Genetic Transformation in Streptococcus pneumoniae: Mutations in σA Bypass the comW Requirement. *Journal of Bacteriology* 196.21 (2014): 3724-3734.

TEACHING EXPERIENCE

University of Illinois at Chicago

Teaching Assistant, Mendelian and Molecular Genetics	**Fall 2013, Spring 2014, Spring 2015, Fall 2015**
Teaching Assistant, Microbiology Laboratory	**Fall 2014**
Teaching Assistant, Cellular Biology Lab	**Fall 2012, Spring 2013**
Teaching Assistant, Cells and Organisms	**Fall 2011, Spring 2012**

PROFESSIONAL ASSOCIATIONS

American Society for Microbiology
Graduate Women In Science (GWIS), Eta Chapter, Chicago
Chicago Area Undergraduate Research Symposium Judge
Positive Thinking – Gram Positive Organisms Seminar Club, Chicago

RESEARCH PRESENTATIONS

Oral Presentations

Tovpeko, Y. and Morrison, D.A. September 2015. Bacteria give each other the best gifts: DNA exchange in *Streptococcus pneumoniae* is mediated by sigma factor switching. Ripon University, Wisconsin. **Invited Speaker.**

Tovpeko, Y. and Morrison, D.A. June 2015. Primary sigma factor mutations bypass the *comW* requirement during competence for genetic transformation in *Streptococcus pneumoniae*. 59.43. 59th Annual Wind River Conference on Prokaryotic Biology, Estes Park, Colorado.

Tovpeko, Y. and Morrison, D.A. August 2014. Mutations in σ^A bypass the ComW requirement for genetic transformation in *S. pneumoniae*. 2014:47. 2014 Molecular Genetics of Bacteria and Phages Meeting, Madison, Wisconsin.

Tovpeko, Y. and Morrison, D.A. June 2014. Regulation of the alternative sigma factor, σ^X, during transformation in *S. pneumoniae*. 58:41. 58th Annual Wind River Conference on Prokaryotic Biology, Estes Park, Colorado.

Tovpeko, Y. Hasan, O., Harrison, G., Morrison, D.A. June 2013. Pro-sigma or Anti-anti-sigma? The Role of the σ^x Partner, ComW, in Regulation of Competence for Transformation in *Streptococcus pneumoniae.*7:37. 7th International Conference on Gram-Positive Microorganisms, Montecatini-Terme, Italy.

Poster Presentations

Tovpeko, Y. and Morrison, D.A. August 2015. Mutations in σ^A bypass the ComW requirement in *S. pneumoniae*. 22:54. 22nd Annual Midwest Microbial Pathogenesis Conference, Chicago, Illinois.

Tovpeko, Y. and Morrison, D.A. August 2015. Mutations in σ^A bypass the ComW requirement for genetic transformation in *S. pneumoniae*. 2015:85. 2015 Molecular Genetics of Bacteria and Phages Meeting, Madison, Wisconsin.

Tovpeko, Y. and Morrison, D.A. September 2014. Mutations in σ^A bypass the ComW requirement for genetic transformation in *S. pneumoniae*. 21:139. 21st Annual Midwest Microbial Pathogenesis Conference, Chicago, Illinois.

Inniss, N., **Tovpeko, Y.**, Bai, J., Morrison, D.A. June 2012. ComX Dependent Competence in *S. pneumoniae*: Characterization of ComW Suppressors. 56:63. Poster given at 56th Annual Wind River Conference for Prokaryotic Biology, Las Vegas, Nevada.

Tovpeko, Y., Lebel, R.R. May 2011. The Basis and Implications of Genetic Counseling in Patient Testing Choice. 2011:259. Talk presented at the Renée Crown University Honors Senior Capstone Presentation Day, Syracuse, New York.

MENTORSHIP EXPERIENCE

Mentor for Undergraduate Research Assistants

Beatrice Go - Honors Capstone Thesis Research, UIC LAS Research Initiative Recipient	**Fall 2014 - Present**
Samim Taraji - Honors Capstone Thesis Research	**Fall 2013 - Spring 2014**
Osamah Hasan - Honors Capstone Thesis Research	**Fall 2012 - Spring 2014**
Gregory Harrison - Summer Undergraduate Researcher	**Summer 2012, Summer 2013**

Mentor for Laboratory Assistants **Fall 2012 - Present**

- Trained and oversaw the employment of undergraduate laboratory assistants for Work Study
- Undergraduates: Anna Do, Om Bhetuwal, Marlin Amy Halder

LEADERSHIP EXPERIENCE

Advisor, Member of Alpha Phi Omega - Co-ed Community Service Fraternity **2007 - Present**

Syracuse University and University of Illinois at Chicago
Staff Advisor: advising president and current members about fraternity affairs and community service
Vice President of Service (December 2009 – May 2010)

Student Peer Advisor and Counselor **Fall 2009 – May 2011**

Syracuse University, Syracuse, New York

- Advised first year students on classes to suit their major requirements and interests

LANGUAGE SKILLS

English (Native), Russian (Fluent), Italian (Moderate).

PERMISSION TO REPRINT