IDENTIFICATION AND CHARACTERIZATION OF DNA RECEPTOR CANDIDATES IN *HAEMOPHILUS INFLUENZAE*

by

SEBASTIAN JAMES MOLNAR
B.Sc., York University, 2001

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES
(Genetics Graduate Program)

THE UNIVERSITY OF BRITISH COLUMBIA

August 2004

ABSTRACT

Haemophilus influenzae is a naturally transformable bacterium with a mechanism for the preferential uptake of conspecific DNA. This mechanism requires the specific recognition of uptake signal sequences (USS) on exogenous DNA by receptors on the bacterial cell surface. The USS-specific DNA receptor in *H. influenzae* is unknown. From the set of known and putative competence genes, candidate USS-receptor genes were identified using bioinformatic approaches. ComE1 was initially identified as the best candidate receptor, since it is orthologous to proteins that bind DNA for uptake in *Bacillus subtilis* and other bacteria. A *comE1* knockout was previously found to have a 10-fold reduction in its transformability, and this retention of partial DNA uptake in Δ*comE1* allowed a direct test for a change in uptake specificity. Thus, to test for a role as the USS-specific DNA-receptor, I used DNA-competition experiments to compare uptake specificity in the Δ*comE1* and wild type strains. A loss of uptake specificity in Δ*comE1* was not detected. Therefore, *comE1* does not encode the USS-specific receptor.

Recently, a putative competence operon (*comNOPQ*) was found to have an upstream competence regulatory element (CRE), which is essential for competence gene expression, and microarray analysis showed that its genes are induced during competence development. The operon contained USS-receptor candidate genes, however, it was not known whether *comNOPQ* genes function in competence. To test this, I constructed a knockout of *comN* and measured its transformability. The transformation frequency of the Δ*comN* strain was 10^5-fold less than that of wild type. Furthermore, I found that DNA binding and uptake were completely eliminated in the Δ*comN*. This suggests that *comN* could be a component of the DNA binding and uptake machinery or that it functions in the assembly of such apparatus. An alternative possibility is that the Δ*comN* strain has polar effects on downstream genes in the *comNOPQ* operon, and *comN* does not function in the competence mechanism. Further work is needed to test these possibilities. However, I have demonstrated that the *comNOPQ* operon is part of the competence regulon and that it contains one or more genes that are essential for the DNA uptake mechanism in *H. influenzae*.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BHI	brain heart infusion (rich culture medium)
cAMP	3',5' cyclic adenosine monophosphate
CFU	colony forming unit
CPM	counts per minute
CRP	cAMP receptor protein
CRE	competence regulatory element
DNA	deoxyribonucleic acid
DNaseI	deoxyribonuclease I
HI#	*Haemophilus influenzae* gene number
IM	inner membrane
kan	kanamycin resistance gene cassette
kan	kanamycin resistance allele
MIV	"M-four", starvation medium for *H. influenzae* competence induction
NAD	nicotinamide adenine dinucleotide
nov	novobiocin resistance allele
OM	outer membrane
PCR	polymerase chain reaction
PS	periplasmic space
sBHI	BHI supplemented with hemin and NAD
SA	signal anchor
SE	standard error of the mean
SEC	secreted
SP	signal peptide
TMH	transmembrane helix
USS	uptake signal sequence

ACKNOWLEDGEMENTS

I want to thank my supervisor, Dr. Rosemary Redfield, for accepting me as a Masters Student in her lab and for providing expert advice and financial support throughout my studies. I also give thanks to the members of my graduate committee, Dr. Redfield, Dr. Sally Otto, and Dr. Phil Hieter for their guidance on my projects and for reading early drafts of this manuscript. For financial support, I thank the Genetics Graduate Program and the Department of Zoology for providing Teaching Assistantships. As well, being a TA for Bio 336 for three consecutive years was a valuable learning experience. Thanks again goes the GGP for hiring me as their website manager. Tremendous gratitude goes out to all members of the Redfield lab for their constructive criticisms, as well as for our lab hikes and beer drinking sessions. In particular, Andrew Cameron has been an excellent office companion and helpful lab mate over the years. And thanks to Dr. Mohammed Bakkali for our many interesting conversations and excursions. I must also thank the Vancouver salsa community for much needed distractions from academia. Most importantly, I am grateful to my family, friends, and colleagues, who have supported me towards my goals – to them, I dedicate this work.

Chapter 1: General Introduction

The work presented here is the combination of two main projects. The findings were written with the intention for publication in the Journal of Bacteriology. Thus, Chapters 2 and 3 are intended to stand-alone. Chapter 4 is a detailed discussion about the results used in chapters 2 and 3 from my bioinformatics studies for predicting competence protein function. The present chapter is included to provide a general background into the main underlying theme of my M.Sc. thesis: understanding the mechanism of DNA uptake in *Haemophilus influenzae*.

The first project focused on identifying the gene or genes that encoded the USS-specific DNA-binding receptor in *H. influenzae*. From a complete list of the CRE regulon genes in *H. influenzae* (Redfield *et al.*, in prep.), candidate USS receptor genes were identified using bioinformatic techniques. The best candidate was *comE1* (HI1008), since it has orthologs known to bind DNA in other naturally competent bacteria (Chapter 2). A previous lab technician, Caixa Ma, had constructed a *comE1* knockout strain (RR3002) and found that deletion of *comE1* had reduced the transformation frequency by only 10-fold when compared to wild type. Hence, the competence phenotype of the Δ*comE1* strain allowed for the direct testing of *comE1* as the USS-specific receptor. My contribution to the project included testing the Δ*comE1* strain for defects in DNA binding and USS-specific DNA uptake. The results from this project are presented in Chapter 2.

In the light of gene expression data (Redfield *et al.*, in prep), I undertook a separate, but related, project from that described in Chapter 2. In Chapter 3, I demonstrate that the putative competence operon (HI0938 to HI0941) contains at least one gene required for DNA uptake. Details of the mutant construction for Chapter 3, is provided as supplementary material in Appendix A. Appendix B contains the raw data generated from the localization predictions for proteins that are encoded by CRE regulated genes, used in chapters 2 and 3, and described in more detail in Chapter 4.

1.1 Natural Competence and Sequence Specific DNA Uptake in *H. influenzae*

Natural competence is the genetically programmed physiological state allowing bacteria to take up exogenous DNA. Natural competence is widespread and is found in numerous Gram negative and Gram positive bacteria (8, 15). A diagram of the DNA uptake mechanism is shown in Figure 1. Several explanations have been given for why competence evolved (1, 8, 18-20, 22), but these remain controversial due in part to a limited understanding of the DNA binding and uptake machinery.

The interesting feature of natural competence in *Haemophilus influenzae* is that this organism preferentially takes up conspecific DNA (8, 17, 23, 24, 27). This preferential uptake mechanism requires the specific recognition of an "uptake signal sequence" (USS) on exogenous DNA by receptors on the competent cell's surface (5, 23, 24). The *H. influenzae* USS consists of a conserved 9-base pair core (5'-AAGTGCGGT), surrounded by an AT-rich 3' flanking region (26). Several other members of the Pasteurellaceae family have the same USS-specific DNA uptake mechanism as competent *H. influenzae*

(1, 30). Species in the Neisseriaceae also have sequence specific DNA uptake, but the neisserial USS (5'-GCCGTCTGAA) is different from the pasteurellacean USS (12, 17). In contrast, most naturally competent bacteria, including *Bacillus subtilis* and other Gram positive species, lack this sequence-specific DNA uptake mechanism and can take up DNA from any source (8). Despite extensive research on USSs and uptake specificity in *Haemophilus* (1, 4, 5, 10, 24, 26) and in *Neisseria* (11, 17, 25), the identities of the USS receptors in the Pasteurellaceae and Neisseriaceae remain unknown.

We expect that the USS-specific receptors are cell surface proteins (5) and any competence proteins that localize to the cell surface during competence are ideal candidates. Therefore, I used protein localization prediction algorithms to identify cell surface competence proteins in *H. influenzae*, and make predictions about possible functions for CRE regulon genes (Chapters 2 and 4). I tested one of the receptor candidates (*comE1*) for a role in USS-specific DNA binding and the results are described in Chapter 2.

1.2 *H. influenzae* Competence Gene Regulation

Although competence genes have been identified in *H. influenzae* (2, 7, 28, 29), the precise functions of many of these during the sequence-specific DNA-uptake process are poorly understood (e.g., the USS-specific DNA receptor is unknown). Determining what genes are involved and how they function during DNA uptake, is crucial for understanding the mechanism and the evolutionary significance of competence.

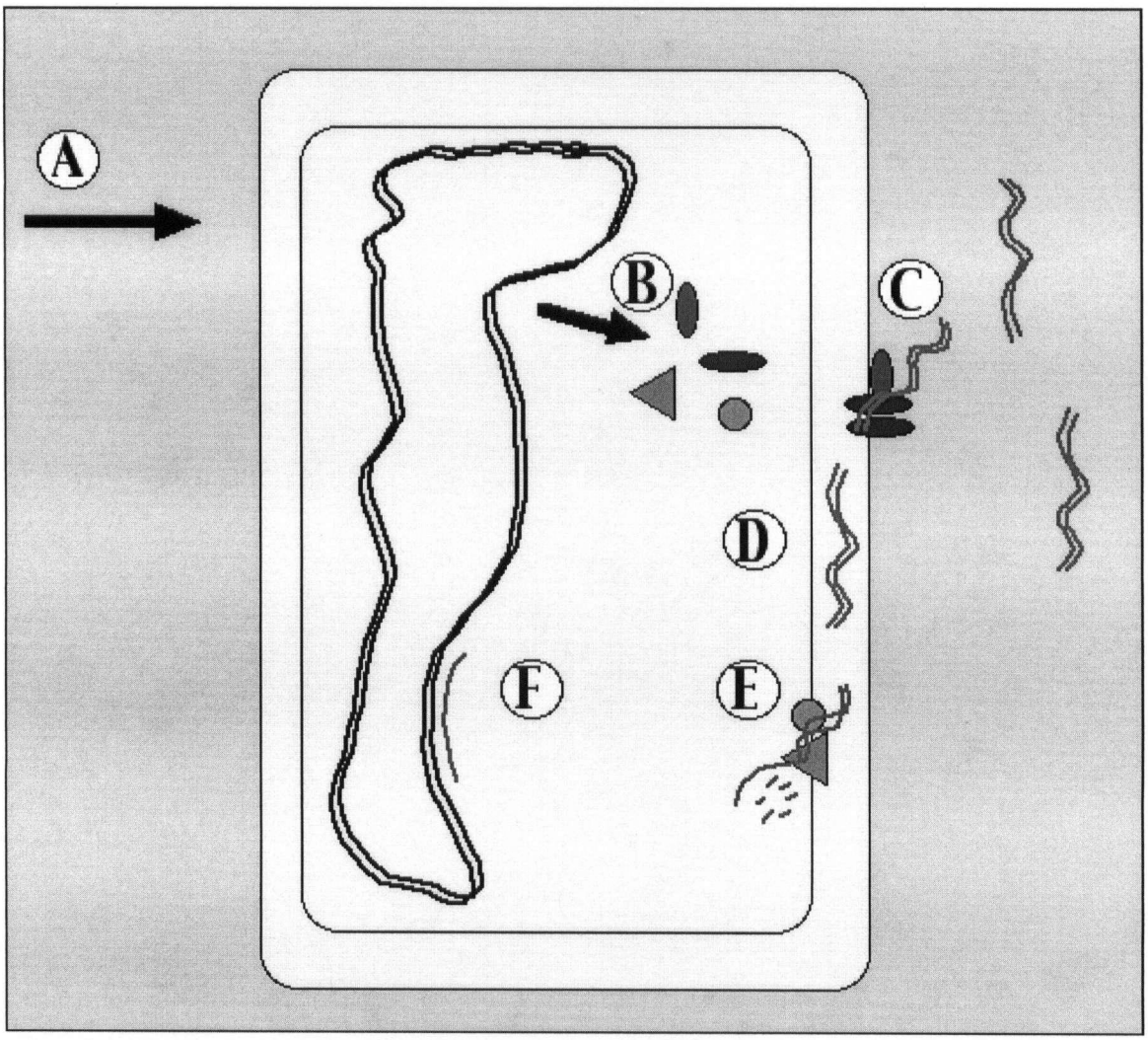


Figure 1.1: A model of competence in *H. influenzae*.

(A) A signal from the environment induces (B) competence gene expression in an *H. influenzae* cell. (C) DNA binding proteins on the cell surface specifically bind to dsDNA containing USSs (uptake signal sequences). (D) The surface bound DNA is then taken up across the outer membrane and into the periplasm. (E) The DNA becomes translocated across the inner membrane, where the 5' strand is degraded by cellular DNases. (F) The intact DNA strand may then continue on to recombine into the bacterial chromosome, or else it too also becomes degraded.

In *Haemophilus influenzae*, genes that are expressed under competence inducing conditions (e.g. by direct transfer to MIV starvation medium) require the presence of a competence regulatory element (CRE) in the promoter regions (9, 14, 16). In total, there are 13 CREs in the *H. influenzae* genome, which specifically regulate nearly 30 genes during competence development (Redfield *et al.*, in prep). Competence genes may be regulated by a transcription factor that specifically recognizes CREs in promoter regions. One possible competence regulator in *H. influenzae* is Sxy (21). The *sxy* gene is known to be essential for competence gene expression in *H. influenzae* (competence is completely abolished in a *sxy* knockout) (21, 31). How Sxy functions in regulation remains to be determined, and is beyond the scope of the present work.

Another factor required for competence regulation is the cAMP-CRP complex (3, 6). Increasing cAMP levels can induce competence, while *cya* (adenylate cyclase) and *crp* mutants are incapable of inducing competence genes (3, 6, 13). Due to the remarkable sequence similarity between CRP binding sites and CREs, it is possible that CREs are CRP-binding sites (16), but further work is required to confirm this hypothesis. The present work does not focus on understanding the competence regulatory mechanism, but uses what is known about competence genes to identify possible USS receptor candidates and to understand the DNA uptake mechanism.

1.3 The identification of a putative competence operon in *H. influenzae*

Gene expression of wild type *H. influenzae* cells under competence inducing conditions was measured using microarrays (Redfield, in prep.). Microarray data analysis revealed

that four genes (HI0938, HI0939, HI0940, HI0941) present in "operon 0938" had similar expression patterns to known competence genes, such as *comA* (Figure-2). Furthermore, upon inspection of the upstream promoter region, operon 0938 was found to have an upstream CRE site (Redfield *et al.*, in prep.). This operon was not previously known to function during natural transformation in *H. influenzae*.

In a separate microarray experiment, gene expression in a *sxy* competence regulatory knockout and in a *cya* (adenylate cyclase) knockout was compared to that of wild type *H. influenzae* (Redfield *et al.*, in prep). If operon 0938 functions in competence, we would expect its expression to be affected by *sxy* and by *cya* (*i.e.*, competence gene expression is reduced in either a *sxy* or a *cya* knockout) (13). Redfield *et al.* (in prep.) found that the expression of operon 0938 was dramatically reduced in the *sxy* knockout as compared to its expression in wild type *H. influenzae*.

The gene expression data, and the presence of the CRE site, strongly suggested that one or more genes in operon 0938 function during competence. Therefore, I have named this operon *comNOPQ*. I originally intended to construct knockouts of all four genes in *comNOPQ* and test them for competence defects. However, due to time constraints, I was only able to construct knockouts of *comN* (HI0938). I tested the Δ*comN* strains for defects in transformation and DNA binding. The details of this project are described in Chapter 3 and Appendix A.

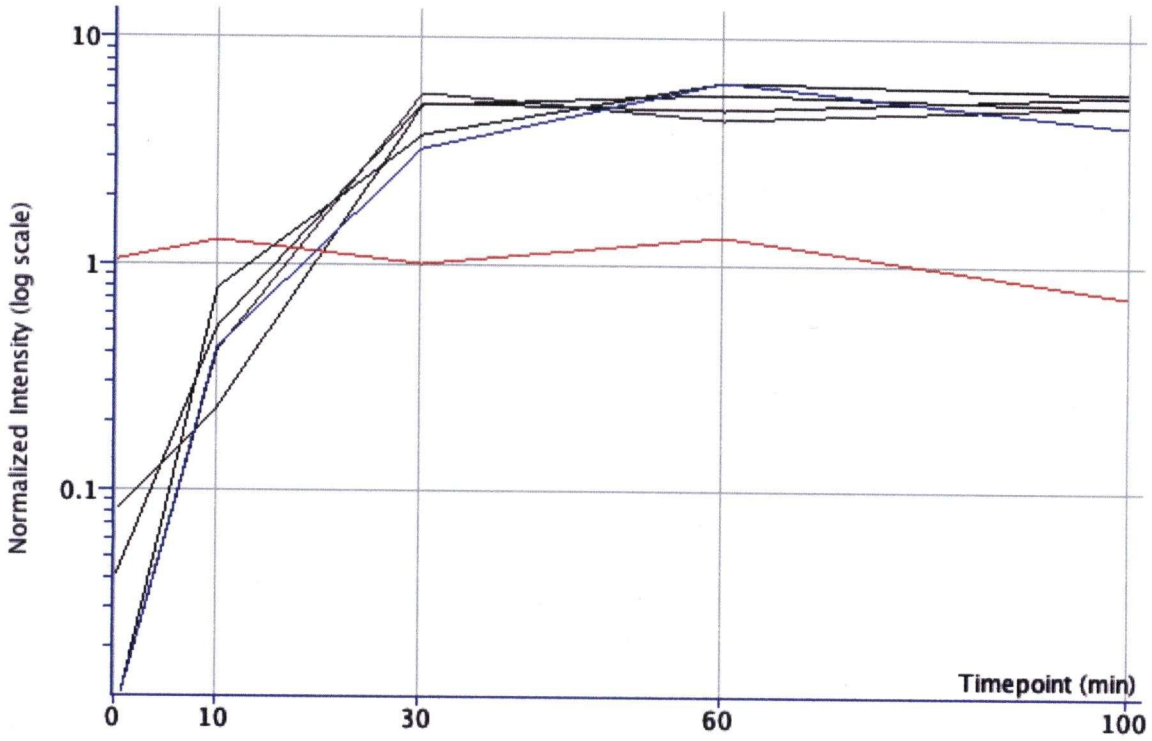


Figure 1.2: Microarray time course data showing wild type gene expression of HI0938, HI0939, HI0940, and HI0941 in comparison with *comA* under competence inducing conditions.

Wild type *H. influenzae* was grown in sBHI and then transferred to MIV medium (time at 0 minutes). During competence induction, samples were taken at regular time points and prepared for microarray analysis (Redfield *et al.*, in prep). Samples in MIV were taken at several time points up to 100 minutes. Genes in operon 0938 (black lines) have similar expression patterns as that of *comA* (blue). To show the expression of a gene that is not affected by competence development, *recB* (red) is included.

1.4 Recapitulation of Objectives

Using the wealth of sequence data available for *H. influenzae*, I have used bioinformatics in an attempt to predict the functions of competence genes and genetic approaches to test these predictions. The following are the research directions followed and the questions that will be addressed in the present study:

A) The *H. influenzae* ComE1 protein is a candidate USS-receptor. Because the Δ*comE1* strain has a reduced, but not abolished, transformation frequency, Δ*comE1* can be directly tested for USS-specific DNA uptake. Does *comE1* encode the USS-specific DNA receptor in competent *H. influenzae* cells? The results of this project are presented in Chapter 2.

B) Do the putative competence genes HI0938 (*comN*), HI0939 (*comO*), HI0940 (*comP*), and HI0941 (*comQ*) function during competence? If so, we expect that knockouts of these genes will affect one or more competence steps (e.g. DNA binding, DNA uptake, translocation, transformation) in *H. influenzae*. To test this, I constructed a knockout of HI0938 (*comN*) and compared it to the wild type strain for defects in transformation and DNA binding. The findings of this study are presented in Chapter 3.

C) Are the predicted cellular locations of competence proteins consistent with the observed phenotypic effects in competence gene knockouts? A discussion of this topic is given in Chapter 4.

1.5 CHAPTER ONE REFERENCES:

1. **Bakkali, M., T. Y. Chen, H. C. Lee, and R. J. Redfield.** 2004. Evolutionary stability of DNA uptake signal sequences in the Pasteurellaceae. Proc Natl Acad Sci U S A **101:**4513-8.
2. **Caster, J. H., E. H. Postel, and S. H. Goodgal.** 1970. Competence mutants: isolation of transformation deficient strains of *Haemophilus influenzae*. Nature **227:**515-7.
3. **Chandler, M. S.** 1992. The gene encoding cAMP receptor protein is required for competence development in *Haemophilus influenzae* Rd. Proc Natl Acad Sci U S A **89:**1626-30.
4. **Danner, D. B., H. O. Smith, and S. A. Narang.** 1982. Construction of DNA recognition sites active in *Haemophilus* transformation. Proc Natl Acad Sci U S A **79:**2393-7.
5. **Deich, R. A., and H. O. Smith.** 1980. Mechanism of homospecific DNA uptake in *Haemophilus influenzae* transformation. Mol Gen Genet **177:**369-74.
6. **Dorocicz, I. R., P. M. Williams, and R. J. Redfield.** 1993. The *Haemophilus influenzae* adenylate cyclase gene: cloning, sequence, and essential role in competence. J Bacteriol **175:**7142-9.
7. **Dougherty, B. A., and H. O. Smith.** 1999. Identification of *Haemophilus influenzae* Rd transformation genes using cassette mutagenesis. Microbiology **145 (Pt 2):**401-9.
8. **Dubnau, D.** 1999. DNA uptake in bacteria. Annu Rev Microbiol **53:**217-44.
9. **Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J. F. Tomb, B. A. Dougherty, J. M. Merrick, and et al.** 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science **269:**496-512.
10. **Goodgal, S. H., and M. A. Mitchell.** 1990. Sequence and uptake specificity of cloned sonicated fragments of *Haemophilus influenzae* DNA. J Bacteriol **172:**5924-8.
11. **Goodman, S. D., and J. J. Scocca.** 1991. Factors influencing the specific interaction of *Neisseria gonorrhoeae* with transforming DNA. J Bacteriol **173:**5921-3.
12. **Goodman, S. D., and J. J. Scocca.** 1988. Identification and arrangement of the DNA sequence recognized in specific transformation of *Neisseria gonorrhoeae*. Proc Natl Acad Sci U S A **85:**6982-6.
13. **Gwinn, M. L., A. E. Stellwagen, N. L. Craig, J. F. Tomb, and H. O. Smith.** 1997. In vitro Tn7 mutagenesis of Haemophilus influenzae Rd and characterization of the role of atpA in transformation. J Bacteriol **179:**7315-20.
14. **Karudapuram, S., and G. J. Barcak.** 1997. The *Haemophilus influenzae* dprABC genes constitute a competence-inducible operon that requires the product of the tfoX (sxy) gene for transcriptional activation. J Bacteriol **179:**4815-20.
15. **Lorenz, M. G., and W. Wackernagel.** 1994. Bacterial gene transfer by natural genetic transformation in the environment. Microbiol Rev **58:**563-602.
16. **Macfadyen, L. P.** 2000. Regulation of competence development in *Haemophilus influenzae*. J Theor Biol **207:**349-59.

17. **Mathis, L. S., and J. J. Scocca.** 1982. *Haemophilus influenzae* and *Neisseria gonorrhoeae* recognize different specificity determinants in the DNA uptake step of genetic transformation. J Gen Microbiol **128 (Pt 5):**1159-61.
18. **Redfield, R. J.** 1988. Evolution of bacterial transformation: is sex with dead cells ever better than no sex at all? Genetics **119:**213-21.
19. **Redfield, R. J.** 1993. Evolution of natural transformation: testing the DNA repair hypothesis in *Bacillus subtilis* and *Haemophilus influenzae*. Genetics **133:**755-61.
20. **Redfield, R. J.** 1993. Genes for breakfast: the have-your-cake-and-eat-it-too of bacterial transformation. J Hered **84:**400-4.
21. **Redfield, R. J.** 1991. sxy-1, a *Haemophilus influenzae* mutation causing greatly enhanced spontaneous competence. J Bacteriol **173:**5612-8.
22. **Redfield, R. J., M. R. Schrag, and A. M. Dean.** 1997. The evolution of bacterial transformation: sex with poor relations. Genetics **146:**27-38.
23. **Scocca, J. J., R. L. Poland, and K. C. Zoon.** 1974. Specificity in deoxyribonucleic acid uptake by transformable Haemophilus influenzae. J Bacteriol **118:**369-73.
24. **Sisco, K. L., and H. O. Smith.** 1979. Sequence-specific DNA uptake in *Haemophilus* transformation. Proc Natl Acad Sci U S A **76:**972-6.
25. **Smith, H. O., M. L. Gwinn, and S. L. Salzberg.** 1999. DNA uptake signal sequences in naturally transformable bacteria. Res Microbiol **150:**603-16.
26. **Smith, H. O., J. F. Tomb, B. A. Dougherty, R. D. Fleischmann, and J. C. Venter.** 1995. Frequency and distribution of DNA uptake signal sequences in the *Haemophilus influenzae* Rd genome. Science **269:**538-40.
27. **Solomon, J. M., and A. D. Grossman.** 1996. Who's competent and when: regulation of natural genetic competence in bacteria. Trends Genet **12:**150-5.
28. **Tomb, J. F., G. J. Barcak, M. S. Chandler, R. J. Redfield, and H. O. Smith.** 1989. Transposon mutagenesis, characterization, and cloning of transformation genes of *Haemophilus influenzae* Rd. J Bacteriol **171:**3796-802.
29. **Tomb, J. F., H. el-Hajj, and H. O. Smith.** 1991. Nucleotide sequence of a cluster of genes involved in the transformation of *Haemophilus influenzae* Rd. Gene **104:**1-10.
30. **Wang, Y., S. D. Goodman, R. J. Redfield, and C. Chen.** 2002. Natural transformation and DNA uptake signal sequences in *Actinobacillus actinomycetemcomitans*. J Bacteriol **184:**3442-9.
31. **Williams, P. M., L. A. Bannister, and R. J. Redfield.** 1994. The *Haemophilus influenzae sxy-1* mutation is in a newly identified gene essential for competence. J Bacteriol **176:**6789-94.

Chapter 2: The Role of a Competence Protein, ComE1, in *Haemophilus influenzae.*

Sebastian J Molnar, Caixa Ma and Rosemary J Redfield

2.1 ABSTRACT

Haemophilus influenzae is a naturally transformable bacterium with a mechanism for the preferential uptake of conspecific DNA. This mechanism requires the recognition of uptake signal sequences (USS) on exogenous DNA by receptors on the bacterial cell surface. The USS-specific DNA receptor in *H. influenzae* has not been identified. The best candidate receptor is ComE1, a homolog of proteins that bind DNA for uptake by *Bacillus subtilis* (ComEA) and other bacteria. ComE1 is part of the CRE competence regulon in *H. influenzae*, and is predicted to localize at, or near, the competent cell surface. We constructed a *comE1* knockout, and found that its transformability was reduced by only 10-fold. This indicates that *comE1* is not essential for DNA uptake by *H. influenzae*. This is in contrast to known ComEA orthologs in other naturally competent species. We have also demonstrated that Δ*comE1* is defective in DNA binding relative to the wild type strain. The retention of partial competence in Δ*comE1* allowed us to test for a change in uptake specificity. To test for a role as the USS-specific DNA receptor, we used DNA-competition experiments to compare uptake specificity in the Δ*comE1* and wild type strains. A loss of uptake specificity in Δ*comE1* was not detected; therefore, *comE1* does not encode the USS-specific receptor. We suggest that ComE1 is a non-essential competence protein that functions to enhance the binding of exogenous DNA in competent *H. influenzae*.

2.2 INTRODUCTION

Natural competence is the genetically programmed physiological state allowing bacteria to efficiently take up exogenous DNA (19). An important step in this process is DNA-binding by receptors at the cell surface, yet the initial interaction between exogenous DNA and the cell is poorly understood. For example, both *Haemophilus influenzae* and *Neisseria gonorrhoeae* have mechanisms for the preferential uptake of conspecific DNA, which require the recognition of uptake signal sequences (USSs) on exogenous DNA by sequence specific receptors (13, 48, 49). However, despite extensive research on uptake specificity in *H. influenzae* (4, 13, 23, 51) and *N. gonorrhoeae* (24, 36), the identities of sequence-specific DNA receptors in the Pasteurellaceae and Neisseriaceae remain unknown. Identifying such DNA receptors may help in answering why uptake specificity exists in certain species, how USSs became over-represented in the genomes of these species, and how uptake specificity might evolve (4). Furthermore, an important problem in understanding DNA uptake may also be resolved with the identification and functional analyses of DNA receptors: how DNA is transported across bacterial membranes (4).

Natural competence is widespread, as it is found in numerous Gram positive and Gram negative bacteria, and models of the process have been described (19). In general, a signal from the environment induces competence gene expression and is followed by the assembly of protein machinery for DNA binding and uptake. Then, DNA taken into the cell is either degraded or incorporated into the bacterial chromosome (19). In the Gram negative bacterium *H. influenzae*, incoming DNA must traverse the outer membrane (DNA uptake) and the inner membrane (DNA translocation) before DNA can enter the

cytoplasm (19). Understanding the mechanism of DNA binding and uptake (i.e., identifying the proteins involved and determining how they work) is the subject of on going study.

Type IV pilus biogenesis (or type II secretion) genes are known to function during DNA uptake in many naturally transformable bacteria (1, 16, 18, 20, 26, 32, 57, 60). Although *H. influenzae* has not been reported as having structures resembling pili, pilin-like genes in this organism are essential for DNA binding and uptake (16). Pilin-like components for competence in *H. influenzae* include the *pilABCD* operon (16) and *comE*, which is related to *pilQ* in nessierial species (17). Thus far, pili have not been directly implicated in the DNA binding or uptake mechanism (37, 60), and the precise roles of these pilin-like components during the uptake process remain to be determined.

Bacterial species that have USS-specific uptake mechanisms also have genomic USSs that are highly over-represented (51). The *H. influenzae* USS consists of a conserved 9-base pair core (5'-AAGTGCGGT) and an AT-rich 3' flanking region (51). In the *H. influenzae* genome, the USS is present in many more copies than randomly expected. In total, there are 1471 USS copies in the *H. influenzae* genome, while only 8 are expected in a genome of comparable size and GC content (50). Several other members of the Pasteurellaceae family (e.g., *Actinobacillus actinomycetemcomitans*, *Pasteurella multocida*) are also known or predicted to have the same USS-specific DNA uptake mechanism as competent *H. influenzae* (4, 56). As with *H. influenzae*, USSs are enriched in the other pasteurellacean genomes (4, 50). Similarly, *N. meningitidis* and *N.*

gonorrhoeae also have USSs in high genomic abundance, but the neisserial USS (5'-GCCGTCTGAA) is different from the pasteurellacean USS (24, 50). The possible evolutionary forces involved in generating USS abundance over time has been argued elsewhere (4) and will not be discussed here.

Most naturally competent bacteria lack the sequence specific DNA uptake mechanism found in the Pasteurellaceae and Neisseriaceae, and instead take up DNA from any source (19, 33). For example, competent cells of *Bacillus subtilis* bind and take up DNA from *Escherichia coli*, *Salmonella typhimurium*, and phage T7 (33, 42, 52). In *B. subtilis*, the DNA receptor is ComEA, which lacks sequence-specific binding and is essential for transformation (45). ComEA orthologs have been detected in both Gram positive and Gram negative bacteria, including *Streptococcus pneumoniae* (CelA) (41), *Thermus thermophilus* and *Acinetobacter* sp. (*comEA*) (20). As with the *B. subtilis comEA*, these orthologs were shown to be essential for transformation (20, 41).

In the Neisseriaceae, *N. meningitidis* and *N. gonorrhoeae* have *comEA* homologs present in multiple dispersed copies (designated *comE1* to *comE4*) (8). This is unlike *comEA* orthologs in other species, which are present only as single genomic copies. Transformation is dramatically reduced in *N. gonorrhoeae* when all four neisserial copies of *comE* are deleted, but only moderately diminished with two or three copies deleted, and not significantly affected by single-copy deletions (8). Similar to the *B. subtilis* ComEA, neisserial ComE binds DNA non-specifically (8). Thus, ComEA homologs appear to have similar functions in Gram-positive and Gram-negative bacteria.

H. influenzae was previously found to contain a single *comEA* homolog (8, 30, 45), *comE1* (TIGR accession no. HI1008). *comE1* has an upstream CRE site characteristic of genes in the competence regulon (30, 35) and is induced 100-fold during competence development (Redfield, in prep.). The presence of the CRE, coupled with the expression data, strongly suggest that *comE1* is a competence gene in *H. influenzae*. To determine if *comE1* affects competence in *H. influenzae*, we constructed a *comE1* knockout and tested it for defects in transformation and DNA binding.

2.3 MATERIALS AND METHODS

Bacterial strains and culture conditions

Strains and plasmids used in this study are listed in Table 2.1. All *H. influenzae* strains are descendants of the original Rd strain (2). The wild type *H. influenzae* strain used is KW20 (58). Standard methods for *H. influenzae* are described by Poje and Redfield (43, 44). *H. influenzae* strains were grown aerobically at 37°C in brain heart infusion broth (BHI, Difco) supplemented with 10μg/ml hemin and 2μg/ml NAD (sBHI) (22). Where appropriate, cultures were streaked or plated on sBHI solidified with 1.2% Bacto Agar (Difco) or MBI Agar. When required, antibiotics were used at 7.0 μg/ml for kanamycin, and 2.5 μg/ml for novobiocin. *Escherichia coli* strains were grown in Luria-Bertani broth at 37°C either with or without antibiotics, as required (47): kanamycin at 10 μg/ml and ampicillin at 100 μg/ml were used.

Table 2.1: Strains and plasmids used in the present study.

Strain or Plasmid	Relevant Genotype	Source or Reference
H. influenzae		
KW20	wild type	(58)
MAP7	*str, kan, nov, nal, spc, vio, stv*	(7)
RR3004	*glpR*::*kan*	this work
RR3003	*comE1 Bcl* I::*kan*	this work
RR3002	Δ*comE1*::*kan*	this work
E. coli		
DH5α	*sup44 recA1*	(27)
GM2123	*dam-*	Noreen Murray
NM554	*sup0 recA*	Noreen Murray
Plasmids		
pGEM-T Easy	*lacZ, amp*	(Promega)
pComE1 (RR1125)	*comE1*	this work
pComE1::Kan (RR1121)	*kan*	this work
pComE1 *Bcl* I::Kan (RR1122; RR1123)	*kan* inserted at *Bcl* I site	this work
pΔComE1::Kan (RR1124)	Δ*comE1*::*kan*	this work

Sequence analyses

All competence gene sequences in *H. influenzae* were obtained from the TIGR database (The Institute for Genomic Research, www.tigr.org). BLAST (3) searches were used on *H. influenzae* competence protein sequences to find significantly similar sequences in other species. Sequence topology analyses used SignalP (39, 40), TMHMM v2.0 (31), PSORT (38), and PSORT-B (21).

E. coli transformation and cloning procedures

Plasmids were transformed into chemically competent *E. coli* using standard procedures (47). Competent *E. coli* cells (DH5α and GM2165) were prepared by treatment with cold 100 ml $CaCl_2$. Transformants were chosen from LB selective plates. For blue/white screening of clones, using vectors pGEM-T Easy or pCR2.1 TOPO-TA, transformed cultures were serially diluted and grown on LB selective plates containing 80 μg/ml X-GAL.

PCR amplification, cloning, and *comE1* mutagenesis

A 2.85 kb fragment flanking HI1008 (*comE1*) was amplified by PCR using the forward and reverse primers 5'-GCGGATCATAGTGGCTGGCAAC-3' and 5'-GCAAGATACCGCTGCGCCCATACCC-3', respectively. The product was cloned into the pGEM-T easy vector system (Promega) to generate the plasmid pComE1. Transposon mutagenesis failed to produce insertions in *comE1* but resulted in a plasmid, pComE1::Tn10kan, carrying a transposon insertion downstream of *comE1*. The plasmid pComE1 contains a unique *Bcl* I site located upstream of *comE1* between the CRE site

and translational start codon. A 1.8 kb kanamycin resistance gene from miniTn10kan cut with *BamHI* was inserted at the *Bcl* I site of pComE1. This generated the plasmid pComE1 *Bcl* I::*kan* with the same orientation of the kanamycin cassette as in pComE1::Tn10kan.

To make a deletion of *comE1*, pComE1 *BclI*::*kan* and pComE1::*kan* were cut with *XhoI* and *PstI*. Two fragments were generated: a 5.7 kb fragment which contained the vector, the 5' sequence of the PCR fragment to the *BclI* site and half of the transposon from pComE1-*BclI*::*kan*; and a 1.6kb fragment which contained the remaining half of the transposon and most of *glpR* from pComE1::*kan*. These two fragments were gel purified using Gene Clean II Kit (Bio101) and then ligated to generate pΔComE1::*kan* which had the whole *comE1* region deleted and included a 45bp deletion from *glpR*. The linearized construct was then transformed into wild type *H. influenzae* to generate the *comE1* knockout (RR3002).

DNA sequencing

DNA sequencing was performed by the University of British Columbia Nucleic Acid-Protein Service Unit, using ABI AmpliTaq DyeDeoxy Terminator cycle-sequencing chemistry. In preparation for sequencing, pComE1::kan was cut with *Xho* I and *Pst* I to yield two fragments each carrying half of the miniTn10kan transposon. A ~6.1 kb fragment was gel purified, and partially sequenced using a primer complementary to the ends of miniTn10kan (5'-CCACCTTAACTTAATGATTTTTACC-3'). The location of the transposon insertion was confirmed at base pair coordinates 1073554 and 1073555.

Competence induction and transformation of *H. influenzae* strains

H. influenzae cultures were grown in sBHI, and competence was induced by the transfer of exponential-phase cells (OD_{600}≈0.25; approximately 10^9 cells/ml) into MIV starvation medium (28, 59). Cells were incubated for 100 minutes in MIV.

The standard procedure for transforming the wild type *H. influenzae* strain KW20 was followed (6): 1ml competent cells were incubated with 1 μg/ml MAP7 DNA for 15 min at 37°C. To stop DNA uptake, DNase I (Boehringer Mannheim) was then added at 1 μg/ml and the mixtures were incubated for an additional 5 min at 37°C. Transformed cultures were serially diluted and plated on selective (novobiocin) and non-selective sBHI plates. Colonies were counted after 24 hours of incubation. Transformation frequencies were then calculated by dividing the antibiotic resistant CFU/ml by the total CFU/ml.

Radioactive labeling of chromosomal DNA

MAP7 chromosomal DNA was radiolabeled with 30 Ci/mmol [methyl-tritium] thymidine 5'-triphosphate ([^{3}H]-dTTP; Amersham) using nick translation reactions (47). DNaseI was calibrated to obtain maximal [^{3}H]-dTTP incorporation. Nick translation reactions were incubated for 2 hours at 15°C, and then 5 minutes at 65°C to inactivate the DNaseI enzyme. Radioactivity levels were measured with a Beckman scintillation counter. Yields were about 30% radiolabel incorporation or higher (e.g. 5.3 x 10^6 CPM/μg from 1.6x10^7 CPM total). Reactions were purified with QIAquick PCR purification kits

(Qiagen). For DNA uptake experiments, radiolabeled MAP7 DNA was diluted to 10^5 CPM/μg with cold MAP7 DNA.

DNA binding and uptake experiments

Competent *H. influenzae* cells were incubated with [^{3}H]-dTTP labeled MAP7 DNA at 10^5 CPM/ml cells. To determine "total DNA bound and taken up" versus "DNA bound", cells incubated with [^{3}H] DNA were split in half for parallel treatments: "DNase" and "no DNase". To stop DNA uptake and to remove bound DNA on cell surfaces, DNase I (Boehringer Mannheim) was added at 1 μg/ml and the mixtures were placed on ice. To determine total DNA bound and taken up, DNase I was not added. Cells were then centrifuged for 1 minute at 13000 rpm, washed twice with MIV, and prepared for scintillation counting. The negative control included incubating non-competent KW20 cells (i.e. log phase cells, at approximately OD_{600}=0.25) with equivalent amounts of [^{3}H] MAP7. The background control used log-phase KW20 cells that had not been incubated with [^{3}H] MAP7.

Testing USS-specificity of competent Δ*comE1*

For DNA competition experiments, competent KW20 and Δ*comE1* cells were prepared and then transformed with different mixtures of MAP7 DNA with either KW20 DNA or *E. coli* DNA. MAP7 DNA (200 ng) was competed against 0 ng, 400 ng, or 4000 ng of KW20 DNA or *E. coli* DH5α DNA per milliliter of MIV competent cells. Transformation frequencies of Δ*comE1* were then determined and compared to that of the wild type.

2.4 RESULTS

Functional predictions of the *H. influenzae comE1* gene.

The *H. influenzae comE1* gene (HI1008) was previously identified as an ortholog of ComEA, the *Bacillus subtilis* DNA receptor (8, 30, 45). Like neisserial ComEA orthologs (8), the *H. influenzae* ComE1 is much smaller than the *B. subtilis* ComEA (*i.e.*, ComE1 has only 112 amino acid residues, while ComEA has 205). ComE1 has sequence similarity primarily to the ComEA C-terminal region for DNA binding, and lacks most of the ComEA N-terminal region (Figure 2.1). In particular, ComE1 lacks the flexible hinge sequence (QQGGG; Figure 2.1) that is proposed to allow a conformational change in ComEA for presenting bound DNA to the uptake machinery (8, 19, 29). Therefore, the only expected role for ComE1 during competence in *H. influenzae* is DNA binding.

Sequences similar to ComEA are present in many different bacterial species and support the notion that parts of the competence mechanism are evolutionarily conserved (19). Most of the ComE1 sequence is similar to the C-terminal region in ComEA, which contains the helix-hairpin-helix DNA-binding domain (8). ComE1 has 47% identity (70% similarity) with the C-terminal domain of *B. subtilis* ComEA. Good homologs of *comE1* are present in all sequenced Pasteurellaceae. For example, *H. influenzae comE1* has 46% identity (69% similarity) to the *Pasteurella multocida* PM1665 and 47% identity (63% similiarity) to the *Haemophilus ducreyi* HD0650. ComE1 homologs were also found in the unfinished genomes of *Actinobacillus actinomycetemcomitans* (www.genome.ou.edu/act_blast.html) and *Haemophilus somnus* (genome.ornl.gov/microbial/hsom/).

The *B. subtilis* ComEA is an integral membrane protein exposed on the surface of competent cells (29), which is important for its role in DNA binding. Similarly, if ComE1 is a DNA binding protein required for the uptake process in competent *H. influenzae*, we expect ComE1 to localize near the cell surface. Therefore, we used available sequence analysis programs to predict the localization site of ComE1. The SignalP algorithm (39, 40) predicted that ComE1 has a signal peptide that is cleaved at residue 22. The signal peptide is an N-terminal sequence used by the general secretory pathway for targeting proteins first to the inner-membrane, where the signal peptide is then cleaved allowing the mature protein to move to the periplasm, the outer membrane, or outside the cell (46). Consistent with this, the transmembrane protein prediction algorithm TMHMM v 2.0 (31) indicated that *H. influenzae* ComE1 is not an inner-membrane protein. Unfortunately PSORT-B, an algorithm that generates a localization probability for the five Gram-negative bacterial cell compartments (*i.e.*, cytoplasm, inner-membrane, periplasm, outer-membrane, outside the cell), could not predict the final localization site of ComE1. However, with the presence of the signal peptide, ComE1 may localize to the periplasm, the outer membrane, or outside the cell.

In *Bacillus subtilis*, ComEA was shown to be required for both DNA binding and for DNA uptake into the cell (29, 45). The Gram-negative orthologs of ComEA, such as the neisserial and pastuerellacean versions, lack the putative flexible motif needed for DNA uptake in Gram positives (8, 20), and are similar only to the DNA binding region in ComEA. Therefore, it seems likely that the Gram-negative ComEA orthologs function only for DNA binding at the cell surface, and are not for DNA uptake.

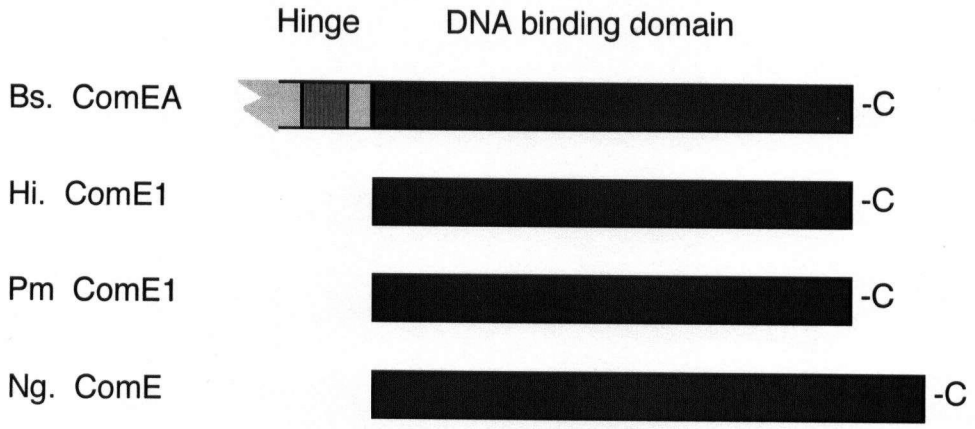


Figure 2.1: Schematic of aligned DNA binding domains in ComEA orthologs.

The C-terminal DNA binding domain in ComEA is a helix-turn-helix motif (8), and is highly conserved among orthologs found in different species. The Gram-negative orthologs all contain N-terminal signal peptides required for protein localization, but multiple alignments indicate there is little sequence conservation in the hydrophobic regions of the transmembrane helices or signal peptides (not shown). The *B. subtilis* ComEA has a hinge motif (QQGGG) that is absent in the Gram-negative orthologs (8). Bs, *B. subtilis* Bsu2555; Hi, *H. influenzae*, HI1008; Pm, *Pasteurella multocida*, PM1665; Ng, *Neisseria gonorrhoeae*, NMA2187.

Construction of a Δ*comE1* strain

To determine whether *comE1* has a role in competence, we constructed a *comE1* knockout and tested it for competence defects. The strategy for generating the plasmids used to construct the Δ*comE1* strain is shown in Figure 2.2. First, a PCR product containing *comE1* and flanking genes was cloned into pGEM-T Easy to generate the pComE1 plasmid. *comE1* is only 339 bp and contains no convenient restriction sites for mutant construction. Transposon mutagenesis failed to produce insertions in *comE1*. Therefore, we fused appropriate fragments from two plasmids, pComE1::Tn10kan and pComE1 *BclI*::kan, to generate the final desired plasmid construct pΔComE1::kan, which contained a kanamycin cassette in place of *comE1* (Figure 2.2).

The pComE1::Tn10kan plasmid was generated by random transposon mutagenesis and contained a miniTn10kan insertion in the 5' end of *glpR*. To construct pComE1 *Bcl*I::kan, a *Bam*HI digested kanamycin cassette was ligated directly into the *Bcl*I restriction site of pComE1. The final construct, pΔComE1::kan, was generated by ligating *Pst*I/*Xho*I double digested fragments from plasmids pComE1 *Bcl*I::kan and pComE1::kan (Figure 2.2). To generate a *comE1* deletion in the chromosome (strain number RR3002), linearized pΔComE1::kan was transformed into wild type *H. influenzae*, and kanamycin resistant transformants were screened for the deleted *comE1* using PCR.

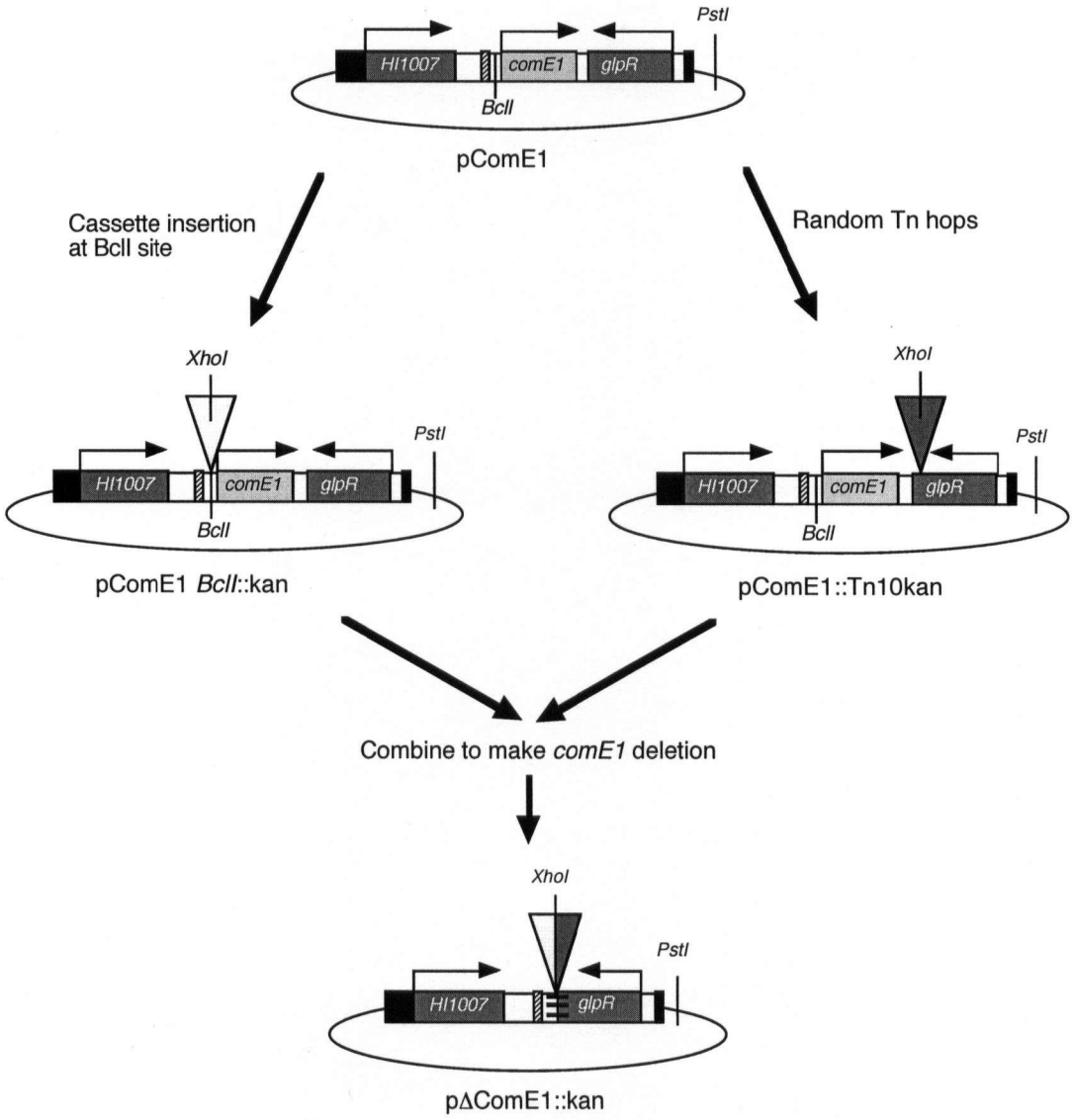


Figure 2.2: Plasmids used to generate the Δ*comE1* strain.

All plasmids used the pGEM-T Easy vector that contained a selectable *amp*^{r} marker. Horizontal arrows show direction of transcription. Hatched boxes represent the CRE site upstream of *comE1*. Black boxes represent truncated flanking genes. The triangles represent the location of the kanamycin cassette or miniTn10kan in the plasmids. The *XhoI* site in pComE1*BclI*::kan and pComE1::Tn10kan was used to generate fragments for fusing the kanamycin cassette from pComE1 *BclI*::kan to the miniTn10kan from pComE1::Tn10kan. Horizontal bars in pΔComE1::kan represents the truncated *glpR*.

Transformability of Δ*comE1*

Defects at any step in the competence mechanism (i.e. competence gene regulation, assembly of uptake machinery, DNA binding, DNA uptake, translocation, or homologous recombination) may cause a reduction in transformation frequencies up to 10^8-fold, whereas differences in uptake of labeled DNA may only be detectable up to 10-fold. Therefore, the most sensitive test for measuring competence in *H. influenzae* strains is the determination of transformation frequencies.

When maximal competence was induced by MIV starvation medium, and competent cells were incubated with saturating amounts of marked DNA, the transformation frequency of Δ*comE1* was only 10-fold less than that of wild type *H. influenzae* (KW20). Using data from five independent experiments, the transformation frequencies were 1.7×10^{-3} (SE ± 0.3×10^{-3}) for KW20, and 1.4×10^{-4} (SE ± 0.3×10^{-4}) for Δ*comE1*. This demonstrates that ComE1 does have a role in competence, but it is not essential for transformation in *H. influenzae*.

Testing growth rates of Δ*comE1* versus KW20

Previously, wild type *comE1* was found to be induced by 100-fold during competence development (Redfield et al., in prep.). To find out whether log phase growth defects were present in Δ*comE1* that affected competence development, it was necessary to do a time course comparison for growth between the Δ*comE1* and KW20 strains. We did not find any significant difference in the doubling time of Δ*comE1* versus the wild type

KW20 strain (data not shown). This indicates that *comE1* does not have a function in log phase growth and supports a specific role for *comE1* in competence.

glpR 5' terminal deletion does not affect competence

Other than growth rate, another possible explanation for any competence phenotype in Δ*comE1* is the effects of adjacent genes. The Δ*comE1* strain (RR3002) includes a 45 bp deletion at the end of *glpR* (H1009). *glpR* is a putative repressor for the glycerol-3-phosphate regulon (TIGR) and is transcribed in the opposite direction as *comE1* (Figure 2.2). *glpR* is not expected to affect competence, since *glpR* expression does not change during competence induction in MIV (Redfield, unpublished). To test for competence defects by the *glpR*, we constructed a chromosomal *glpR* mutant, by transforming linear pComE1::Tn10kan into the wild type strain, and compared its transformability with wild type. We found no significant difference in either the growth or transformation frequency of the *glpR* strain (RR3004) when compared to wild type *H. influenzae* (data not shown). Therefore, the *glpR* deletion in the Δ*comE1* strain is not responsible for the reduction in transformability.

Testing DNA Binding and Uptake in Δ*comE1*

H. influenzae ComE1 has sequence similarity to DNA binding and uptake proteins in other species; therefore, we expected ComE1 to have a similar role in *H. influenzae*. Thus, we investigated whether ComE1 is directly involved in DNA binding and uptake by determining how much DNA entered competent Δ*comE1* cells compared to the wild type.

To measure the total amount of cell-associated DNA (i.e. DNA bound plus taken up), competent cells were incubated for 15 minutes with radiolabeled MAP7 DNA and were not treated with DNase (adding DNase removes surface bound DNA). From two independent experiments, the average total amounts of radiolabeled DNA associated with pellets of KW20 and Δ*comE1* cells were 7586 CPM (SE ± 418 CPM) and 1104 CPM (SE ± 185 CPM), respectively. Therefore, Δ*comE1* binds and takes up DNA seven times less efficiently than wild type. This is consistent with the 10-fold reduction in transformation frequency of Δ*comE1* versus KW20.

ComE1 is not likely to be involved in the post-uptake steps (e.g. translocation or transformation), because the amount of radiolabeled DNA that enters Δ*comE1* cells is much less than wild type. Otherwise, equivalent levels of radiolabeled DNA bound and taken up by wild type and Δ*comE1* cells would be observed. Thus, we conclude that ComE1 is involved at the DNA binding and uptake steps during competence.

Δ*comE1* directly affects DNA binding

Having discovered that our Δ*comE1* strain was defective at the DNA binding and uptake steps, we attempted to determine at which of these steps ComE1 is directly involved. We compared the amount of radiolabeled DNA bound and taken up by DNase versus non-DNase treated competent cells. DNase added to cells not only degrades exogenous free DNA, but also removes surface bound DNA. Therefore, the difference between non-DNase versus DNase treated cells should equal the amount of surface bound DNA.

The Δ*comE1* DNA binding and uptake phenotype can be explained by either of two possible scenarios. If deleting *comE1* results in weakened uptake machinery, but retains wild type DNA binding, we expect that "no DNase" treated cells would have much higher associated radiolabeled DNA compared to "DNase" treated cells. Alternatively, if Δ*comE1* has wild type uptake machinery, but weak DNA binding, then we would expect nearly equivalent and low values of radiolabeled DNA associated with either DNase treated or non-treated cells. In which case, Δ*comE1* would display weak DNA binding and have wild type DNA uptake machinery.

In one trial, competent Δ*comE1* cells bound 10 times less radiolabeled DNA than did KW20 cells (i.e. 134 CPM versus 1524 CPM, respectively). However, in separate trials, this precise ratio was not reproducible. For example, the ratio of [^{3}H] DNA bound by KW20 versus [^{3}H] DNA bound by Δ*comE1*, ranged from 5:1 (10840 CPM vs. 2332 CPM) to 50:1 (2364 CPM vs. 42 CPM) when cells were incubated with [^{3}H] DNA for different amounts of time. Thus, an accurate determination of total DNA bound versus DNA taken up could not be made. This is likely due to the poor sensitivity in detecting surface bound DNA. However, Δ*comE1* consistently bound less DNA than wild type. Thus, our data suggests that Δ*comE1* has a defective, but weak DNA-binding phenotype.

Testing the USS-specificity of Δ*comE1*

Since ComE1 is the only *H. influenzae* competence protein with known DNA-binding homologs in other species, it is a candidate for the USS receptor. We thus set out to test whether the Δ*comE1* strain is defective for USS-specific DNA binding.

To determine if ComE1 affects USS-specificity during competence, we competed MAP7 DNA (using the selectable marker, *nov*^{R}) against unmarked KW20 DNA or *E. coli* DNA, at different concentrations during transformation of competent cells (Figure 2.3). Using DNA competition during transformation is the classical method used to identify uptake specificity in *H. influenzae* (48, 49). When marked *H. influenzae* DNA is competed against increasing amounts of unmarked *H. influenzae* DNA, we expect the transformation frequency to decrease in wild type cells. In contrast, when marked *H. influenzae* DNA is competed against increasing amounts of unlabeled *E. coli* DNA, we do not expect the transformation frequency to change in wild type cells due to the uptake specificity for conspecific DNA. If Δ*comE1* is defective in USS-specific DNA binding, we expect the number of transformants will decrease when MAP7 DNA is competed against increasing amounts of either umarked KW20 DNA or *E. coli* DNA.

Although transformation frequencies in Δ*comE1* are 10-fold lower than KW20, both strains exhibited similar trends during DNA competition experiments. The MAP7 DNA concentration was maintained at 200 ng/ml of competent cells for each sample, while the competing unmarked DNA was used at 0 ng, 400 ng, or 4000 ng/ml of cells. In both competent Δ*comE1* and KW20 cells, the transformation frequencies decreased when MAP7 DNA was competed against KW20 DNA, but did not significantly decrease when MAP7 DNA was competed against *E. coli* DNA (Figure 2.3). This is due to the preference for USS containing DNA during DNA binding and uptake in *H. influenzae*. Thus, ComE1 is not the USS-specific DNA receptor in *H. influenzae*.

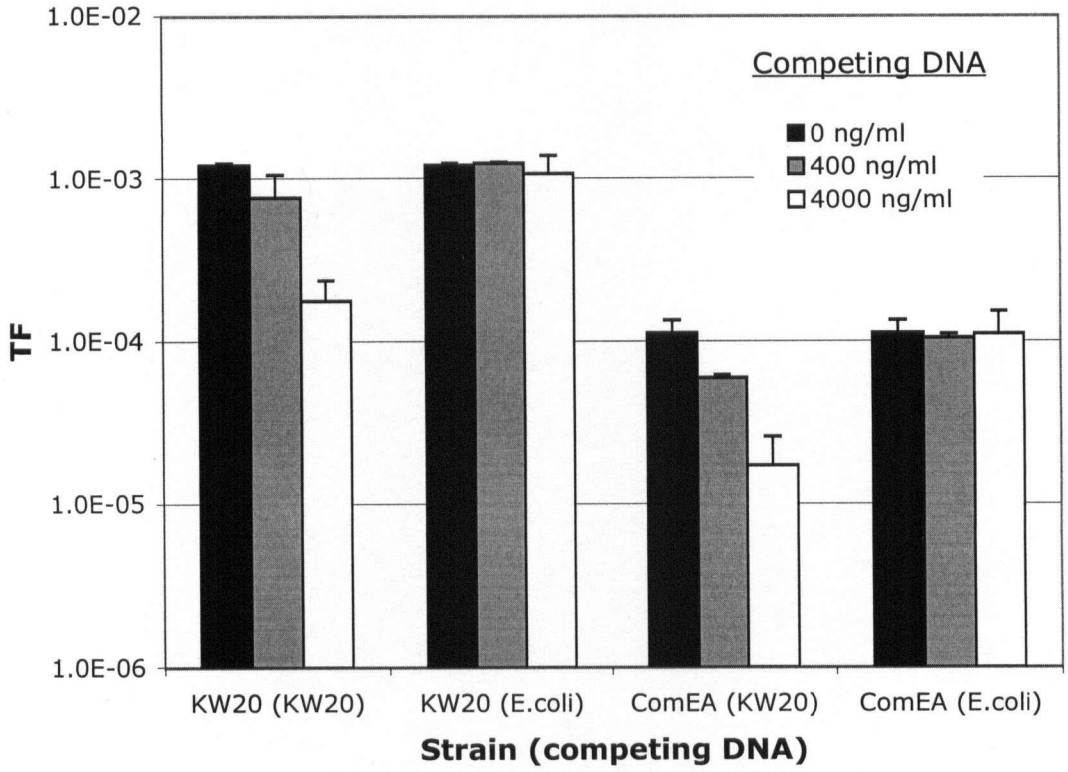


Figure 2.3: Competition for DNA uptake.

Competent KW20 and Δ*comE1* cells were incubated with MAP7 DNA (200 ng/ml cells) and with the indicated amount of competing KW20 DNA or *E.coli* DNA. The figure legend indicates the amount of the unmarked competing DNA from either KW20 or *E. coli*. The results represent the mean transformation frequencies from two independent experiments, and the error bars represent the standard error.

Identification of candidate USS-receptor genes in *H. influenzae*

Because *comE1* does not encode the USS-receptor, we attempted to identify other candidate USS-receptor genes from known or putative competence genes in *H. influenzae* (*i.e.*, genes that are expressed under competence inducing conditions and have CRE sites in their promoter regions; Redfield et al., in prep.). The ideal candidate for a USS-receptor is a protein that localizes to the surface of competent cells.

Using the protein sequence analysis algorithms SignalP, TMHMM, and PSORT-B, we found six competence genes that encode proteins with cleavable signal peptides, indicating they are sent to the periplasm or the outer membrane. These are HI0436 (*comD*), HI0435 (*comE*), HI0940 and HI0941 (hypothetical proteins), HI1008 (*comE1*) and HI1182/83 (DNA ligase). Another competence protein (ComB) was initially predicted to have a signal peptide, but upon further inspection it was found to have a signal anchor (*i.e.*, non-cleavable signal peptide) or transmembrane helix. The HI0938, HI0939, and *pilA* genes are predicted to encode pilin or pilin-like proteins (Chapter 3) that may localize to the periplasm, outer membrane, or outside the cell. Three competence genes (*rec2*, *pilC*, and *pilD*) encode polytopic inner-membrane proteins with 11, 3, and 6 predicted transmembrane helices, respectively. The remaining thirteen known or putative competence proteins were identified as being cytoplasmic (Chapter 4).

The protein localization predictions were used to guide us in identifying the most likely USS-receptor candidates. The *H. influenzae comE1* gene (HI1008) is the only candidate with orthologs that are known to have DNA-binding properties in other competent

bacteria. However, we have demonstrated that ComE1 is not the USS-specific DNA receptor. Another candidate gene, *comE* (HI0435, which is located in a different operon than *comE1*, HI1008), was predicted to encode an outer-membrane protein related to the PulD family of secretins. A knockout of *comE* is known to eliminate transformation and DNA binding (54). Its relatedness to the PulD family of secretins, including proteins such as PilQ in neisserial species (10, 11), suggests that ComE may have a function in either the transport of DNA across the outer-membrane (e.g. it may act as a channel for incoming DNA), or in assembly of the DNA-binding and uptake machinery (it may act as a channel to transport the uptake machinery proteins). It is possible that ComE functions as the USS-specific DNA binding protein, and our lab is currently testing ComE for a role as the USS-receptor (Redfield, personal communication). The HI1182/83 gene encodes a periplasmic DNA ligase (9), and therefore is not expected to bind DNA on the cell surface. Thus, the most likely candidates for a USS-receptor gene in competent *H. influenzae* are HI0436 (*comD*), HI0438 (*comE*), HI0299 (*pilA*) and HI0938 to HI0941.

2.5 DISCUSSION

In our attempts to identify the USS-specific receptor for competence in *Haemophilus influenzae*, we determined that one candidate, ComE1, is a DNA binding protein. ComE1 had previously been identified due to its sequence similarity to known DNA binding proteins, such as ComEA in *Bacillus subtilis*, CelA in *Streptococcus pneumoniae* and ComE in *Neisseria* sp. (8, 29, 41). The *comEA* homologs are required for transformation in these species. In contrast, we found that *H. influenzae* ComE1 is not

essential for natural transformation, because the Δ*comE1* strain has only a 10-fold reduced transformation frequency compared to the wild type strain.

We have presented evidence that a Δ*comE1* strain is defective in DNA binding relative to the wild type. Furthermore, a *comE1* knockout retains USS-specificity during natural transformation. Thus, *comE1* does not encode the USS-specific receptor. The present study complements the work by Chen and Gotshlich (8), which demonstrated that the nessieral ComEA orthologs directly bind DNA, but do so non-specifically. It seems likely that *comEA* orthologs encode non-specific DNA-binding receptors in many naturally competent bacteria, as such homologs are found in numerous species (8). It was suggested that species with sequence-specific DNA uptake mechanisms have two different types of DNA receptors: one for non-specific, and another for sequence-specific, DNA-binding (15). This is consistent with the transformation phenotype observed in our Δ*comE1* strain (*i.e.*, Δ*comE1* still retains uptake specificity).

Since ComE1 is not essential for natural transformation in *H. influenzae*, and it is not the USS-receptor, it may play a role as an accessory DNA-binding protein during competence. That is, ComE1 may assist a USS-specific receptor in binding DNA for the efficient uptake into competent *H. influenzae* cells. It is possible that the non-specific receptor binds DNA first, and then the USS-receptor scans bound DNA for the USS core (8). In contrast, ComEA homologs in *Neisseria* spp. were suggested to have an essential role in binding DNA after USS specific binding takes place (8). This indicates that not

only are the USS-sequences different between the Pasteurellaceae and Neisseriaceae, but that their mechanisms of DNA binding are different also.

An important question in cell biology is how macromolecules are transported across cell membranes. The USS-specific DNA binding and uptake mechanisms pose interesting examples. Based on sequence features of the USS core and its AT-rich flanking regions, it is tempting to speculate how DNA receptors bind exogenous DNA and mediate DNA uptake across the outer-membrane. Competent *H. influenzae* cells bind double-stranded, rather than single-stranded, DNA (55), and incoming dsDNA does not require a free end for the binding and uptake steps (5). At the cell surface scale, dsDNA is a stiff molecule, and therefore dsDNA transport across the outer-membrane may require local strand separation to allowing kinking (4). The DNA receptor could facilitate the melting of the AT-rich region surrounding the USS core (4, 12, 34). Since DNA bending is a common feature of many DNA binding proteins (14, 25), a USS-receptor could initiate DNA bending in the AT-rich region flanking the USS core for DNA transport across the membrane (34). Another possible mechanism for DNA transport across the outer-membrane is that the DNA receptor nicks one strand of bound dsDNA prior to uptake, causing a local flexible region in bound DNA. Although it was previously shown that dsDNA remains intact immediately after the uptake step (53, 55), the periplasmic DNA ligase (HI1182/83) that is part of the CRE regulon (Redfield, pers. comm.) could repair the nick before the DNA is translocated into the cytoplasm. Identifying the USS-specific receptor will be useful to test such speculations.

2.6 REFERENCES:

1. **Aas, F. E., M. Wolfgang, S. Frye, S. Dunham, C. Lovold, and M. Koomey.** 2002. Competence for natural transformation in *Neisseria gonorrhoeae*: components of DNA binding and uptake linked to type IV pilus expression. Mol Microbiol **46:**749-60.
2. **Alexander, H. E., and G. Leidy.** 1951. Determination of inherited traits of H. influenzae by desoxyribonucleic acid fractions isolated from type-specific cells. J Exp Med **93:**345-59.
3. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J Mol Biol **215:**403-10.
4. **Bakkali, M., T. Y. Chen, H. C. Lee, and R. J. Redfield.** 2004. Evolutionary stability of DNA uptake signal sequences in the Pasteurellaceae. Proc Natl Acad Sci U S A **101:**4513-8.
5. **Barany, F., M. E. Kahn, and H. O. Smith.** 1983. Directional transport and integration of donor DNA in Haemophilus influenzae transformation. Proc Natl Acad Sci U S A **80:**7274-8.
6. **Barcak, G. J., M. S. Chandler, R. J. Redfield, and J. F. Tomb.** 1991. Genetic systems in *Haemophilus influenzae*. Methods Enzymol **204:**321-42.
7. **Catlin, B. W., J. W. Bendler, 3rd, and S. H. Goodgal.** 1972. The type b capsulation locus of *Haemophilus influenzae*: map location and size. J Gen Microbiol **70:**411-22.
8. **Chen, I., and E. C. Gotschlich.** 2001. ComE, a competence protein from *Neisseria gonorrhoeae* with DNA-binding activity. J Bacteriol **183:**3160-8.
9. **Cheng, C., and S. Shuman.** 1997. Characterization of an ATP-dependent DNA ligase encoded by Haemophilus influenzae. Nucleic Acids Res **25:**1369-74.
10. **Collins, R. F., L. Davidsen, J. P. Derrick, R. C. Ford, and T. Tonjum.** 2001. Analysis of the PilQ secretin from Neisseria meningitidis by transmission electron microscopy reveals a dodecameric quaternary structure. J Bacteriol **183:**3825-32.
11. **Collins, R. F., R. C. Ford, A. Kitmitto, R. O. Olsen, T. Tonjum, and J. P. Derrick.** 2003. Three-dimensional structure of the Neisseria meningitidis secretin PilQ determined from negative-stain transmission electron microscopy. J Bacteriol **185:**2611-7.
12. **Danner, D. B., H. O. Smith, and S. A. Narang.** 1982. Construction of DNA recognition sites active in *Haemophilus* transformation. Proc Natl Acad Sci U S A **79:**2393-7.
13. **Deich, R. A., and H. O. Smith.** 1980. Mechanism of homospecific DNA uptake in *Haemophilus influenzae* transformation. Mol Gen Genet **177:**369-74.
14. **Dickerson, R. E.** 1998. DNA bending: the prevalence of kinkiness and the virtues of normality. Nucleic Acids Res **26:**1906-26.
15. **Doran, J. L., and W. J. Page.** 1983. Heat sensitivity of Azotobacter vinelandii genetic transformation. J Bacteriol **155:**159-68.
16. **Dougherty, B. A., and H. O. Smith.** 1999. Identification of *Haemophilus influenzae* Rd transformation genes using cassette mutagenesis. Microbiology **145 (Pt 2):**401-9.

17. **Drake, S. L., and M. Koomey.** 1995. The product of the *pilQ* gene is essential for the biogenesis of type IV pili in *Neisseria gonorrhoeae*. Mol Microbiol **18:**975-86.
18. **Dubnau, D.** 1997. Binding and transport of transforming DNA by *Bacillus subtilis*: the role of type-IV pilin-like proteins--a review. Gene **192:**191-8.
19. **Dubnau, D.** 1999. DNA uptake in bacteria. Annu Rev Microbiol **53:**217-44.
20. **Friedrich, A., T. Hartsch, and B. Averhoff.** 2001. Natural transformation in mesophilic and thermophilic bacteria: identification and characterization of novel, closely related competence genes in Acinetobacter sp. strain BD413 and Thermus thermophilus HB27. Appl Environ Microbiol **67:**3140-8.
21. **Gardy, J. L., C. Spencer, K. Wang, M. Ester, G. E. Tusnady, I. Simon, S. Hua, K. deFays, C. Lambert, K. Nakai, and F. S. Brinkman.** 2003. PSORT-B: Improving protein subcellular localization prediction for Gram-negative bacteria. Nucleic Acids Res **31:**3613-7.
22. **Goodgal, S. H., and R. M. Herriott.** 1961. Studies on transformations of Hemophilus influenzae. I. Competence. J Gen Physiol **44:**1201-27.
23. **Goodgal, S. H., and M. A. Mitchell.** 1990. Sequence and uptake specificity of cloned sonicated fragments of *Haemophilus influenzae* DNA. J Bacteriol **172:**5924-8.
24. **Goodman, S. D., and J. J. Scocca.** 1988. Identification and arrangement of the DNA sequence recognized in specific transformation of *Neisseria gonorrhoeae*. Proc Natl Acad Sci U S A **85:**6982-6.
25. **Goodman, S. D., N. J. Velten, Q. Gao, S. Robinson, and A. M. Segall.** 1999. In vitro selection of integration host factor binding sites. J Bacteriol **181:**3246-55.
26. **Graupner, S., V. Frey, R. Hashemi, M. G. Lorenz, G. Brandes, and W. Wackernagel.** 2000. Type IV pilus genes pilA and pilC of Pseudomonas stutzeri are required for natural genetic transformation, and pilA can be replaced by corresponding genes from nontransformable species. J Bacteriol **182:**2184-90.
27. **Hanahan, D.** 1983. Studies on transformation of Escherichia coli with plasmids. J Mol Biol **166:**557-80.
28. **Herriott, R. M., E. M. Meyer, and M. Vogt.** 1970. Defined nongrowth media for stage II development of competence in *Haemophilus influenzae*. J Bacteriol **101:**517-24.
29. **Inamine, G. S., and D. Dubnau.** 1995. ComEA, a *Bacillus subtilis* integral membrane protein required for genetic transformation, is needed for both DNA binding and transport. J Bacteriol **177:**3045-51.
30. **Karudapuram, S., and G. J. Barcak.** 1997. The *Haemophilus influenzae* dprABC genes constitute a competence-inducible operon that requires the product of the tfoX (sxy) gene for transcriptional activation. J Bacteriol **179:**4815-20.
31. **Krogh, A., B. Larsson, G. von Heijne, and E. L. Sonnhammer.** 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol **305:**567-80.
32. **Link, C., S. Eickernjager, D. Porstendorfer, and B. Averhoff.** 1998. Identification and characterization of a novel competence gene, comC, required for DNA binding and uptake in Acinetobacter sp. strain BD413. J Bacteriol **180:**1592-5.

33. **Lorenz, M. G., and W. Wackernagel.** 1994. Bacterial gene transfer by natural genetic transformation in the environment. Microbiol Rev **58:**563-602.
34. **Lundin, M., J. O. Nehlin, and H. Ronne.** 1994. Importance of a flanking AT-rich region in target site recognition by the GC box-binding zinc finger protein MIG1. Mol Cell Biol **14:**1979-85.
35. **Macfadyen, L. P.** 2000. Regulation of competence development in *Haemophilus influenzae*. J Theor Biol **207:**349-59.
36. **Mathis, L. S., and J. J. Scocca.** 1982. *Haemophilus influenzae* and *Neisseria gonorrhoeae* recognize different specificity determinants in the DNA uptake step of genetic transformation. J Gen Microbiol **128 (Pt 5):**1159-61.
37. **Mathis, L. S., and J. J. Scocca.** 1984. On the role of pili in transformation of Neisseria gonorrhoeae. J Gen Microbiol **130 (Pt 12):**3165-73.
38. **Nakai, K., and P. Horton.** 1999. PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. Trends Biochem Sci **24:**34-6.
39. **Nielsen, H., S. Brunak, and G. von Heijne.** 1999. Machine learning approaches for the prediction of signal peptides and other protein sorting signals. Protein Eng **12:**3-9.
40. **Nielsen, H., J. Engelbrecht, S. Brunak, and G. von Heijne.** 1997. A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Int J Neural Syst **8:**581-99.
41. **Pestova, E. V., and D. A. Morrison.** 1998. Isolation and characterization of three Streptococcus pneumoniae transformation-specific loci by use of a lacZ reporter insertion vector. J Bacteriol **180:**2701-10.
42. **Piechowska, M., A. Soltyk, and D. Shugar.** 1975. Fate of heterologous deoxyribonucleic acid in *Bacillus subtilis*. J Bacteriol **122:**610-22.
43. **Poje, G., and R. J. Redfield.** 2003. General methods for culturing *Haemophilus influenzae*. Methods Mol Med **71:**51-6.
44. **Poje, G., and R. J. Redfield.** 2003. Transformation of *Haemophilus influenzae*. Methods Mol Med **71:**57-70.
45. **Provvedi, R., and D. Dubnau.** 1999. ComEA is a DNA receptor for transformation of competent *Bacillus subtilis*. Mol Microbiol **31:**271-80.
46. **Pugsley, A. P.** 1993. The complete general secretory pathway in gram-negative bacteria. Microbiol Rev **57:**50-108.
47. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular Cloning, A Laboratory Manual, 2nd Edt. Cold Spring Harbour Laboratory Press, USA.
48. **Scocca, J. J., R. L. Poland, and K. C. Zoon.** 1974. Specificity in deoxyribonucleic acid uptake by transformable *Haemophilus influenzae*. J Bacteriol **118:**369-73.
49. **Sisco, K. L., and H. O. Smith.** 1979. Sequence-specific DNA uptake in *Haemophilus* transformation. Proc Natl Acad Sci U S A **76:**972-6.
50. **Smith, H. O., M. L. Gwinn, and S. L. Salzberg.** 1999. DNA uptake signal sequences in naturally transformable bacteria. Res Microbiol **150:**603-16.
51. **Smith, H. O., J. F. Tomb, B. A. Dougherty, R. D. Fleischmann, and J. C. Venter.** 1995. Frequency and distribution of DNA uptake signal sequences in the *Haemophilus influenzae* Rd genome. Science **269:**538-40.

52. **Soltyk, A., D. Shugar, and M. Piechowska.** 1975. Heterologous deoxyribonucleic acid uptake and complexing with cellular constituents in competent *Bacillus subtilis*. J Bacteriol **124:**1429-38.
53. **Stuy, J. H.** 1965. Fate of transforming DNA in the *Haemophilus influenzae* transformation system. J Mol Biol **13:**554-70.
54. **Tomb, J. F., H. el-Hajj, and H. O. Smith.** 1991. Nucleotide sequence of a cluster of genes involved in the transformation of *Haemophilus influenzae* Rd. Gene **104:**1-10.
55. **Voll, M. J., and S. H. Goodgal.** 1961. Recombination during transformation in Hemophilus influenzae. Proc Natl Acad Sci U S A **47:**505-12.
56. **Wang, Y., S. D. Goodman, R. J. Redfield, and C. Chen.** 2002. Natural transformation and DNA uptake signal sequences in *Actinobacillus actinomycetemcomitans*. J Bacteriol **184:**3442-9.
57. **Wang, Y., W. Shi, W. Chen, and C. Chen.** 2003. Type IV pilus gene homologs pilABCD are required for natural transformation in *Actinobacillus actinomycetemcomitans*. Gene **312:**249-55.
58. **Wilcox, K. W., and H. O. Smith.** 1975. Isolation and characterization of mutants of *Haemophilus influenzae* deficient in an adenosine 5'-triphosphate-dependent deoxyribonuclease activity. J Bacteriol **122:**443-53.
59. **Williams, P., W. L. Hung, and R. J. Redfield.** 1996. Cell transfer by filtration: evaluation of protocols for transformational competence. FEMS Microbiol Lett **137:**183-7.
60. **Wolfgang, M., P. Lauer, H. S. Park, L. Brossay, J. Hebert, and M. Koomey.** 1998. PilT mutations lead to simultaneous defects in competence for natural transformation and twitching motility in piliated *Neisseria gonorrhoeae*. Mol Microbiol **29:**321-30.

Chapter 3: Identification and Characterization of a New Competence Operon in *Haemophilus influenzae*.

Sebastian J Molnar and Rosemary J Redfield

3.1 ABSTRACT

Haemophilus influenzae is a naturally competent Gram-negative bacterium with a mechanism for the preferential uptake of conspecific DNA. Recently, microarray analyses revealed the presence of four genes (*comN*, *comO*, *comP* and *comQ*), not previously associated with competence, that had similar expression patterns to known competence genes such as *comA*. The putative competence operon (*comNOPQ*) has an upstream competence regulatory element (CRE), which is essential for competence gene expression. The presence of the CRE site coupled with the gene expression data strongly suggested that one or more of the *comNOPQ* genes are involved in competence. To test this, we constructed a knockout of *comN* (HI0938) and measured its transformability. The transformation frequency of the Δ*comN* strain was near 10^{-8}, while that of wild type was 10^{-3}. To determine what other competence steps were affected in the Δ*comN* strain, we incubated competent cells with radiolabeled chromosomal DNA and compared the amount of DNA bound and taken up by Δ*comN* and wild type cells. We found that DNA binding and uptake was completely eliminated in Δ*comN*. This suggests that *comN* is a component of the DNA binding and uptake machinery or that it functions in the assembly of such apparatus. An alternative possibility is that the Δ*comN* strain has polar effects on downstream genes in the *comNOPQ* operon, and *comN* does not function in the competence mechanism. Further work is needed to determine whether *comN* is part of

the competence machinery. However, we have demonstrated that the *comNOPQ* operon is part of the competence regulon and it contains genes that are essential for the DNA uptake mechanism in *H. influenzae*.

3.2 INTRODUCTION

Natural competence is the genetically programmed physiological state allowing bacteria to efficiently take up exogenous DNA. Although found in numerous Gram negative and Gram positive bacteria, the competence mechanism differs in certain features among different species (19, 33). For example, some bacterial species, such as *Haemophilus influenzae* and *Neisseria gonorrhoeae*, have a competence mechanism for the preferential uptake of conspecific DNA (19, 36, 51, 52, 54). Other species, such as *Bacillus subtilis* and other Gram positives, are able to take up DNA from any source (19).

The competence process occurs in multiple steps. In general, a signal from the environment induces competence gene expression. Thereafter, protein machinery is assembled for DNA binding and uptake, and DNA taken into the cell is either degraded or incorporated into the bacterial chromosome through natural transformation (19). There are, however, many aspects of the competence mechanism that are poorly understood. For example, the DNA receptor in *H. influenzae* is not known and the precise roles of several competence genes during DNA uptake are unclear. Furthermore, the evolutionary function of natural competence is controversial (6, 19, 44-47). Hence, determining what genes are involved and how they function during DNA uptake is

crucial for understanding the mechanism and the evolutionary significance of competence.

Several *H. influenzae* competence genes have homologs in other species, and this may assist in the determination of competence gene function, as well as in understanding the evolutionary reasons for DNA uptake. For example, genes in the *H. influenzae comABCDEF* operon have been detected in *E. coli* and may function in taking up DNA for nutritional benefits (21). In addition, some *H. influenzae* competence genes are related to type IV pilus-biogenesis genes in *N. gonorrhoeae* and *Pseudomonas aeruginosa* (21).

Type IV pilus biogenesis genes are known to function in the competence mechanisms of many naturally transformable bacteria (1, 16, 18, 22, 26, 32, 60, 63). Although *H. influenzae* has not been reported to have a pilus structure, pilin-like genes in this organism are essential components of the DNA binding and uptake machinery (17). For example, the *H. influenzae comE* gene is related to the well-characterized neisserial gene, *pilQ* (17, 19, 58). *H. influenzae* also contains a *pilABCD* operon that is essential for competence (16). Thus far, the precise roles of these pilins and pilin-like components during the uptake process remain to be determined for *H. influenzae*.

Recently, gene expression of wild type *Haemophilus influenzae* cells under competence inducing conditions was measured using microarrays (Redfield et al., in prep.). Microarray data analysis revealed that four genes (HI0938, HI0939, HI0940, and

HI0941) had similar expression patterns to known competence genes, such as *comA*. Upon inspection of the upstream promoter region, this four gene cluster was found to have a CRE site required for competence gene expression (30, 34). The gene expression data, and the presence of the CRE site, strongly suggested that one or more of these four genes function during competence. Therefore, we named this four-gene cluster, operon *comNOPQ*. In the present work, we constructed a *comN* (HI0938) knockout and tested it for defects in transformation and DNA uptake.

3.3 METHODS and MATERIALS

Bacterial strains and culture conditions

Strains and plasmids used in this study are listed in Table 3.1. All *H. influenzae* strains are descendents of the original Rd strain (2). The wild type *H. influenzae* strain used is KW20 (61). Standard methods for *H. influenzae* are described by Poje and Redfield (40, 41). *H. influenzae* strains were grown aerobically at 37°C in brain heart infusion broth (BHI, Difco) supplemented with 10μg/ml hemin and 2μg/ml NAD (sBHI) (24). Where appropriate, cultures were streaked or plated on sBHI solidified with 1.2% Bacto Agar (Difco) or MBI Agar. When required, antibiotics were used at 7.0 μg/ml for kanamycin and 2.5 μg/ml for novobiocin. *Escherichia coli* strains were grown in Luria-Bertani broth at 37°C with antibiotics as required (49): kanamycin at 10 μg/ml and ampicillin at 100 μg/ml were used.

Table 3.1: Strains and plasmids used in this study.

Strain or Plasmid	Relevant Genotype	Source or Reference
H. influenzae		
KW20	wild type	(61)
MAP7	*str, kan, nov, nal, spc, vio, stv*	(10)
RR938	Δ*comN*::*kan*	(this study)
RR1172	Δ*comN*::*kan*	(this study)
RR1173	Δ*comN*::*kan*	(this study)
RR3002	Δ*comE1*::*kan*	(Molnar et al., in prep)
RR3005	Δ*purR*::*kan*	(Kennith Chan)
E.coli		
DH5α	*sup44 recA1*	(27)
TOP10	*recA1*	Invitrogen
RR921	pWJC3, *kan*	(11)
RR1160	pLEFT	(this study)
RR1161	pRIGHT	(this study)
RR1162	pKAN	(this study)
RR1163	pLEFT_KAN	(this study)
RR1164	pLKR-45::Δ*comN*	(this study)
RR1165	pLKR-46::Δ*comN*	(this study)
Plasmids		
pGEM-T Easy	*lacZ, amp*	Promega
pCR 2.1-TOPO	*lacZ, amp, kan*	Invitrogen
pWJC3	*kan*	(11)
pLEFT	*comNOPQ* CRE site	(this study)
pRIGHT	*comO, comP*	(this study)
pKAN	*kan*	(this study)
pLEFT_KAN	LEFT, *kan*	(this study)
pLKR::Δ*comN*	LKR::Δ*comN*, *kan*	(this study)

Table 3.2: Primers for constructing the *comN* knockout

Primer	Sequence*
P1_PstI	5'- **CTGCAG**TTCGTTGATGCGCAGATAAAGT
P2_BamHI	5'- **GGATCC**TTCCACTAATGTCATACCTTTCTGC
P3	5'- TTTGATAGTGCTTGTAATTAATGATGAA
P4	5'- GCACCTTGATTATCTCCCTTTTTAG
PC1	5'- GATTTTTCTACCCATTCATTTAT
PC2	5'- TGCCCTTTTAATAATACTTTCA
K1_BamHI	5'- **GGATCC**GGGGGGGAAAGCCACGTTTGT
K2	5'- AGCCGCCGTCCCGTCAAGT

*engineered 5'-end restriction sites for primers are indicated in bold.

Sequence analyses

All competence gene sequences in *H. influenzae* were obtained from the TIGR database (The Institute for Genomic Research, www.tigr.org). BLAST (3) searches were used on *H. influenzae* competence protein sequences to find significantly similar sequences in other species. To make predictions about cellular localization of CRE regulated competence proteins, sequence topology analyses were done using SignalP (37, 38), TMHMM v2.0 (31), and PSORT-B (23).

E. coli transformation and cloning procedures

Plasmids were transformed into chemically competent *E. coli* using standard procedures (49). Competent *E. coli* cells (DH5α and GM2165) were prepared by treatment with cold 100 ml $CaCl_2$. Transformants were chosen from LB selective plates. For blue/white screening of clones (e.g. using pGEM-T Easy or with pCR 2.1-TOPO), transformed cultures were serially diluted and grown on LB selective plates containing 80 μg/ml X-GAL.

PCR amplification and cloning

The primers used in this study are listed in Table 3.2. All primers were designed using GeneTool (BioTools Inc.) and NetPrimer (Premier Biosoft International), and were synthesized by Sigma-Genosys Custom Oligos. Expected sizes amplicons were confirmed using Amplify 1.2 (20). Primers P1_*Pst*I and P2_*Bam*HI were used to amplify the upstream flanking region (LEFT) of the *comN* start codon; P3 and P4 are for the downstream region (RIGHT) including 16 bases from the 3' end of *comN*; PC1 and PC2

were used to amplify the *comN* gene starting at 55 bp upstream of the start codon and ending at 27 bp downstream of the stop codon, respectively; and K1_ *Bam*HI and K2 primers are for amplifying the kanamycin resistance cassette (KAN) from pWJC3. The primers P1_*PstI*, P2_ *Bam*HI, and K1_ *Bam*HI were designed to include synthetic restriction sites near the 5' ends (Table 3.2) for use in the knockout procedure.

PCR was performed on a Minicycler (MJ Research Inc.) using the following protocol: denaturation for 5 minutes at 94°C, followed by 30 to 35 cycles of denaturation for 1 minute at 94°C, annealing for 1 minute at 55°C and polymerization at 72°C for 1 minute per kilobase of the expected amplicon, with a final polymerization step at 72°C for 5 minutes. All PCR products were separated and visualized by gel electrophoresis, and purified with QIAquick gel extraction kits (Qiagen) when required.

Restriction digests and Ligation Reactions

The PCR products LEFT and RIGHT were ligated directly into pGEM-T Easy, and KAN into pCR 2.1 TOPO-TA, following the kit procedures by Promega and Invitrogen, respectively. Restriction enzymes (*Bam*HI, *Pst*I, *Spe*I, *Aat*II, *Sac*I, and *Ssp*I) were used to cut out inserts for further ligation reactions or to check for correct insert orientations. Positions of restriction sites were determined with DNA Strider 1.1 (35). Restriction digests were incubated for 1 to 2 hours at 37°C with the recommended buffers for each enzyme, and checked for complete digestion in 1.5% agarose gels. Before further ligation, all inserts and vectors were gel purified using QIAquick gel extraction kits (Qiagen). Ligation reactions were incubated overnight at 4°C with T4 DNA ligase (New

England Biolabs) and 1x T4 Ligase Buffer. To increase the chances of obtaining the desired clones, several insert:vector DNA ratios were used (e.g. at least 2:1 for insert:vector amounts) in different ligation reactions for direct transformation into competent *E. coli*.

Competence induction and transformation of *H. influenzae* strains

H. influenzae cultures were grown in sBHI and competence was induced by the transfer of exponential-phase cells (OD_{600}≈0.25; approximately 10^9 cells/ml) into MIV starvation medium (29, 62). To induce maximal competence levels, cells were incubated in MIV for 100 minutes.

The standard procedure for transforming the wild type *H. influenzae* strain KW20 was followed (7): 1 ml competent cells were incubated with 1 μg/ml MAP7 DNA for 15 min at 37°C. To stop DNA uptake, DNase I (Boehringer Mannheim) was then added at 1 μg/ml and the mixtures were incubated for an additional 5 min at 37°C. Transformed cultures were serially diluted and plated on selective (novobiocin) and non-selective sBHI plates. Colonies were counted after plates were incubated at 37°C for at least 24 hours. Transformation frequencies were then calculated by dividing the antibiotic resistant CFU/ml by the total CFU/ml.

Radioactive labeling of chromosomal DNA

MAP7 chromosomal DNA was radiolabeled with 30 Ci/mmol [methyl-tritium] thymidine 5'-triphosphate (Amersham) using nick translation reactions (49). DNaseI was calibrated

to obtain maximal [^{3}H]-dTTP incorporation. Nick translation reactions were incubated for 2 hours at 15°C, and then 5 minutes at 65°C to inactivate the DNaseI enzyme. Radioactivity levels were measured with a Beckman scintillation counter. Yields were obtained at about 30% radiolabel incorporation or higher (e.g., 5.3 x 10^6 CPM/μg from 1.6x10^7 CPM total). Reactions were purified with QIAquick PCR purification kits (Qiagen). For DNA uptake experiments, radiolabeled MAP7 DNA was diluted to 10^5 CPM/μg with cold MAP7 DNA.

DNA binding and uptake experiments

Competent *H. influenzae* cells were incubated with 1 μg of radiolabeled MAP7 DNA at 10^5 CPM/ml cells. To determine total DNA bound and taken up versus DNA bound, half of the volume of cells incubated with [^{3}H] DNA was used for DNase versus no DNase treatments in parallel. To stop DNA uptake and to remove bound DNA on cell surfaces, DNase I (Boehringer Mannheim) was added at 1 μg/ml and the mixtures were placed on ice. To determine total DNA bound and taken up, DNase I was not added. Cells were then centrifuged for 1 minute at 13000 rpm, washed twice with MIV to remove excess [^{3}H] MAP7, and prepared for scintillation counting. The negative control included incubating non-competent KW20 cells (i.e. log phase cells, at approximately OD_{600}=0.25) with equivalent amounts of [^{3}H] MAP7. The background control used log-phase KW20 cells that had not been incubated with [^{3}H] MAP7.

RESULTS

Plasmid construction to generate Δ*comN*

To determine if *comN* has a role in competence, we constructed a *comN* knockout and tested its transformability. The overall knockout procedure is shown in Figure 3.1. In steps 1, 2, and 3, PCR was used to amplify the segments LEFT and RIGHT from KW20 chromosomal DNA (the 1 kb upstream and downstream flanking regions of the *comN* gene, respectively), and KAN (the kanamycin resistance cassette) from pWJC3. The PCR products were gel purified and ligated into the appropriate vectors (LEFT and RIGHT each into pGEM-T Easy, and KAN into pTOPO) to make the best use of the multiple cloning sites. Ligation reactions were then cloned into *E. coli*, and transformants were selected for using ampicillin LB plates.

Colonies that potentially contained the pLEFT, pRIGHT, and pKAN plasmids were chosen based on blue-white screening (Figure 3.1, steps 1, 2, and 3). Clones were then checked with colony PCR using primers corresponding to the appropriate inserts (i.e. primers P1 and P2 for LEFT; P3 and P4 for RIGHT; and K1 and K2 for KAN). Once candidate clones for plasmids pLEFT, pRIGHT, and pKAN were obtained, restriction digests with *PstI*, *SspI*, and *BamHI*, respectively, were used to check for the correct insert orientations. Clones with the appropriate pLEFT, pRIGHT, and pKAN plasmids were given the strain numbers RR1160, RR1161, and RR1162, respectively.

To construct pLEFT_KAN (pLK), plasmids pLEFT and pKAN were cut with *BamHI* and fragments were gel purified, and the KAN insert was ligated into pLEFT at the *BamHI*

site (Figure 3.1, step 4). Ligations were subsequently transformed into *E. coli*. Colony PCR was used to screen for kanamycin resistant colonies with the LEFT_KAN construct. The orientation of the KAN insert in pLK was determined with PCR (using P1 and K2 primers), since no useful restriction sites were found. Of eleven colonies tested, five were found to contain KAN in the desired orientation. The pLK clone used in step 5 (Figure 3.1) was given the strain number RR1163.

To generate the pLEFT_KAN_RIGHT (pLKR) plasmid, plasmids pLK and pRIGHT were *PstI* digested, and the LK fragment was ligated into the *PstI* site of the pRIGHT vector (Figure-1, step 5). The ligation was transformed directly into competent *E. coli*. Colony PCR was then used to screen clones for the LKR construct and, in parallel, to determine the LK insert orientation in pLKR clones. Correct orientation of the LK insert was confirmed using *PstI* and *SspI* digests on pLKR. Plasmids from two pLKR clones (RR1164 and RR1165) were used to generate *comN* knockouts (Figure 3.1, step 6).

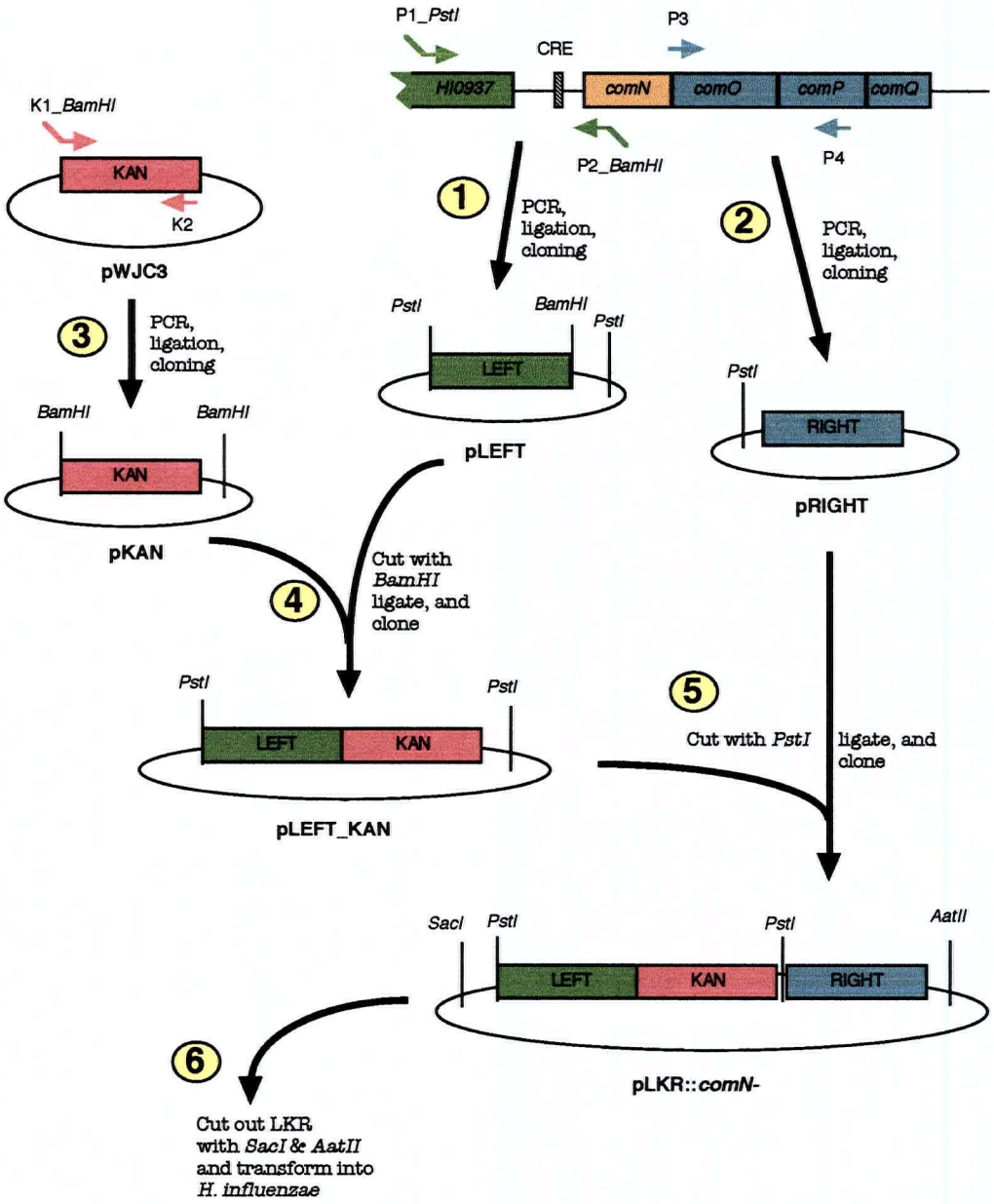


Figure 3.1: Overall scheme for *comN* knockout construction

In steps 1, 2, and 3, PCR was used to amplify the LEFT and RIGHT segments from KW20 chromosomal DNA, and KAN from pWJC3. PCR products were then gel purified and ligated into the vectors pGEM-T Easy (for LEFT and RIGHT) and pTOPO (KAN). Subsequent steps used minipreps, restriction digests, ligation reactions, and cloning (steps 4 & 5) to engineer the pLKR::ΔHI0938 construct. The 3.3 kb LKR construct was cut out of pLKR::ΔHI0938 with *SacI* and *AatII*, and then transformed into competent wild type *H. influenzae* to generate the HI0938 knockout (step 6). CRE, Competence Regulatory Element; LKR, the 3.3 kb LEFT_KAN_RIGHT construct.

Construction of Δ*comN* strains

Once the pLKR plasmid was obtained, it was used to generate the Δ*comN* strain (Figure 3.1, step 6). The 3.3 kb LKR linear construct was cut out of pLKR::Δ*comN* with *SacI* and *AatII*, and then transformed into competent wild type *H. influenzae*. In this transformation step, serial dilutions of digested pLKR::Δ*comN* were used to obtain kanamycin resistant colonies, because higher amounts of DNA increases the likelihood of generating transformants with tandem repeats or other arrangements in the *H. influenzae* chromosome due to the presence of a periplasmic ligase (Poje and Redfield, 2002). Colonies were chosen for screening from plates containing the lowest colony numbers (e.g., clones obtained from transformations using higher DNA dilutions). The positive control was the transformation of wild type competent cells with MAP7 DNA, with subsequent selection on *nov* and *kan* plates. Negative controls were untransformed competent cells (no DNA control) plated on *nov*, *kan*, and *amp* plates. The positive and negative controls gave results as expected: competent wild type KW20 cells had the typical 10^{-3} transformation frequency, and untransformed cells did not grow on antibiotic plates.

Although several of the resulting kanamycin resistant colonies from these transformations on *kan* plates appeared to have normal wild type colony sizes, many others were much smaller than wild type colonies grown on plain plates. Additional controls were included as a check for colony growth of putative Δ*comN* clones on *kan* plates. The Δ*purR* (RR3005) and Δ*comE1* (RR3008) strains were streaked on plain and *kan* plates, since both carry the same pWJC3-derived kanamycin cassette as pLKR::Δ*comN*. After one

day of incubation, colonies from Δ*purR* and Δ*comE1* on *kan* plates had similar sizes to wild type colonies grown on plain plates. However, the putative Δ*comN* microcolonies took two days of incubation on *kan* plates to become visible to the unaided eye. On plain plates, putative Δ*comN* colonies were similar in size to the wild type, Δ*purR* and Δ*comE1* strains after 24 hours of incubation. Reasons for the difference in colony growth on *kan* versus plain plates are unclear.

For rapid screening and to increase the chance of isolating Δ*comN* strains, putative Δ*comN* microcolonies were pooled for colony PCR analysis with K1 and K2 primers. From seven microcolony pools tested, five were identified as containing the KAN segment. From one pool, ten microcolonies were chosen for colony PCR to confirm the presence of KAN. Positive controls included amplifying KAN from the pLKR plasmid and from a *purR* colony. The negative control used a wild type *H. influenzae* colony for the DNA template source and had no amplification in PCR. Of the ten colonies screened, only one had the KAN fragment (strain RR1172). To check for the replacement of *comN* with KAN in the putative Δ*comN* clone, different primer sets were used in parallel for colony PCR (i.e. primers P1/P2 for LEFT; K1/K2, KAN; P3/P4, RIGHT; P1/K2, LEFT_KAN; K1/P4, KAN_RIGHT, and P1/P4 for the full length LKR).

One explanation for the small colony sizes in Δ*comN* isolates is that the pLKR plasmid had a defective *kan* gene; however, this does not seem likely, as there were no colony size differences on *kan* plates of *E. coli* clones harboring the pLKR plasmids. As an additional control, a different construct containing the same kan^R cassette (Δ*pilB*::kan^R

plasmid and chromosomal DNA, obtained from Minghui Yang, unpublished) was transformed into competent KW20 – normal size colonies and microcolonies were still observed. This indicates that the *kan*^{R} cassette might not be properly expressed in the microcolony transformants. Another possibility is that the *comN* gene affects growth on *kan*. This is not expected, because *comN* is not expressed during exponential growth. A third possibility is that the Δ*comN* strain, named RR1172, carries a mutation unlinked to the *comN* deletion that affects growth on *kan*.

To confirm that strain RR1172 was *H. influenzae* and to determine whether the small colony morphology was not due to separate mutations linked to the *comN* deletion, a backcross was done using RR1172 chromosomal DNA to transform KW20 cells. This backcross generated approximately twice as many normal sized *kan*^{R} colonies (when compared to wild type on plain plates) as microcolonies. This indicates that the small colony morphology on kanamycin plates observed in RR1172 may be an independent genetic defect from the *comN* replacement with KAN. To test this, six large colonies were chosen for colony PCR using primer sets PC1/PC2 and K1/K2 (data not shown). All six colonies had the KAN insert (997 bp), however PC1/PC2 amplifications failed to detect any band (1219 bp expected), which is similar to PCR results obtained for the RR1172 strain. One of these clones, a backcross strain, was given the named RR1173.

Due to amplification difficulties with some primer sets on colonies or purified DNA from strain RR1172, an additional transformation experiment was done to generate replicate Δ*comN* strains. A similar mix of colony sizes was observed in separate attempts (*i.e.*, on

kanamycin selective plates, numerous microcolonies and several large colonies were present). From this independent transformation experiment, four additional clones containing the KAN fragment were isolated. PCR with different primer sets (PC1/PC2, P1/K2, and K1/P4) was used with purified DNA to determine the genotypes of these four clones. Only one of the four putative Δ*comN* strains (RR938) had the desired amplification for all primer sets used. However, PCR did reveal that the other three clones were likely to have a replacement of *comN* gene with KAN. Only one primer set failed for these strains (i.e. different primer sets failed for each clone), while the remaining primer sets indicated that *comN* had been replaced. All Δ*comN* strains (RR1172, RR1173, RR938, and three other clones) were checked for competence defects.

Testing for Transformation Defects in Δ*comN* Strains.

The simplest and most sensitive method for detecting competence defects in *H. influenzae* strains is by measuring the transformation frequencies of competent cells. Therefore, we compared the transformation frequencies of Δ*comN*::*kan*^{R} clones (RR1172, RR1173, and RR938) to that of wild type.

In the positive control, wild type *H. influenzae* had a transformation frequency of 10^{-3} as expected. For all Δ*comN*::*kan*^{R} clones, colonies appeared normal on plain plates. On *nov* plates, no colonies were present on undiluted plates for RR1172 or RR1173, while only one colony was observed for RR938. Three other *comN* strains obtained independently from the original microcolony pools (P3-1, P4-1, and P4-2) were similarly tested for transformation defects; these had three or fewer colonies on undiluted *nov* plates. The

transformation frequencies for Δ*comN* strains were reduced by 10^5-fold when compared to wild type (Figure 3.2). This demonstrates that the *comNOPQ* operon contains genes required for competence in *H. influenzae*.

Testing growth rates of Δ*comN* versus KW20

To rule out whether *comN* has a role during log phase growth, which may affect competence development, we compared the growth of Δ*comN* and wild type strains. The Δ*comN* strains which grew as microcolonies on *kan* plates grew similar to wild type cells when grown on plain plates. No significant difference was found in the growth rates of Δ*comN*::*kan*^{R} strains grown in plain broth as compared to wild type (data not shown). Thus, the transformation phenotype in Δ*comN* strains is not due to defects in log phase growth. The results are consistent with a specific role for *comN* in competence.

Testing for DNA binding and uptake defects in Δ*comN* strains.

The second assay for detecting competence defects determines whether or not any DNA is bound or taken up by the cells. Thus, we incubated competent cells with [^{3}H] MAP7 chromosomal DNA and measured the total amount of cell-associated DNA (DNA bound and taken up by cells).

The competent wild type cells used as controls took up about 7% of the labeled DNA (approximately 11000 cpm in total). Non-competent or log phase *H. influenzae* cells were used for a negative control. For Δ*comN* strain RR1172, essentially no ^{3}H-DNA was bound or taken into the cells (Figure 3.3). In total the counts were 75 cpm for RR1172,

which was not significantly different from the negative control (29 cpm). These results demonstrate that the *comNOPQ* operon is essential for natural competence in *H. influenzae*.

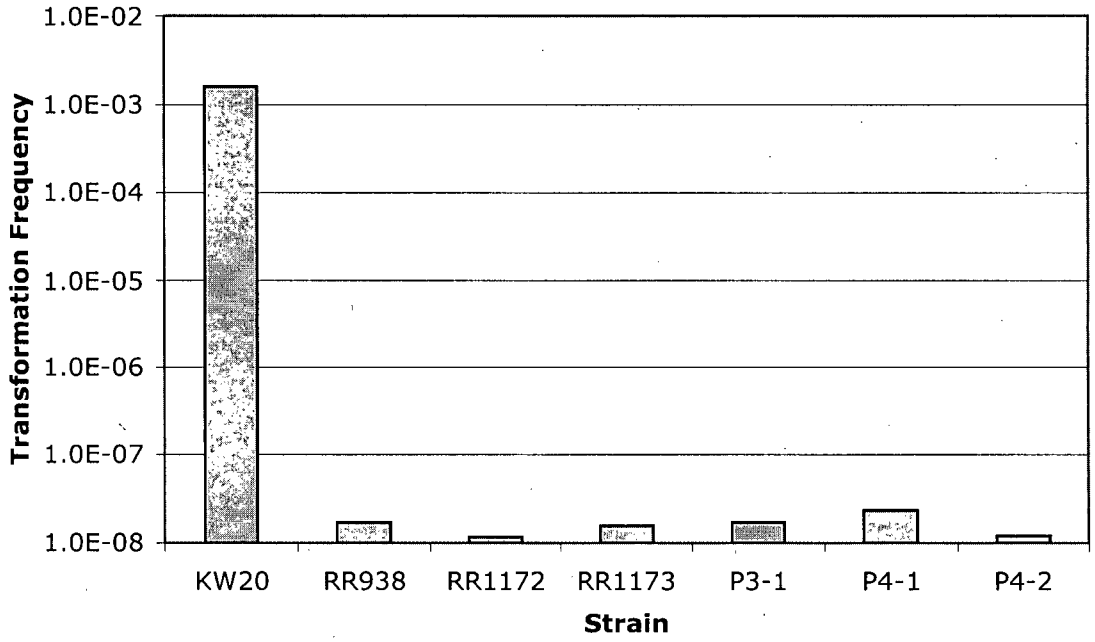


Figure 3.2: Transformation frequencies of Δ*comN* strains versus KW20.

Since no colonies were present on any *nov* plates for strains RR938, RR1172, and RR1173, the transformation frequencies (TF) for these strains were calculated assuming a single colony had been present on undiluted plates. Thus, TFs for RR938, RR1172, and RR1173 are upper detection limits. For strains P3-1, P4-1, and P4-2, the TFs are the actual TFs (2, 3, and 1 colonies were present on undiluted *nov* plates).

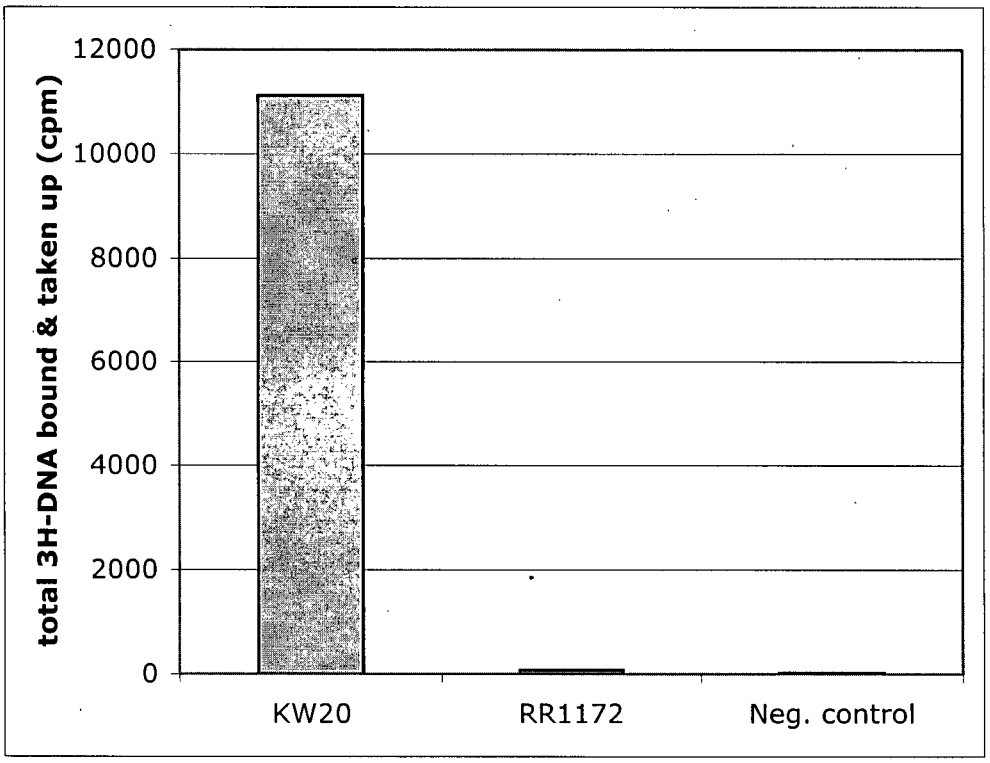


Figure 3.3: DNA bound and taken up by Δ*comN*.

Competent cells were incubated with ^{3}H MAP7 DNA. DNaseI was not added so that total cell-associated DNA (DNA bound and taken up) could be measured. The positive control used competent KW20 cells, and the negative control used non-competent KW20 (cells in log-phase growth).

Localization predictions of ComNOPQ proteins

To predict possible functions of the *comNOPQ* genes, two bioinformatics approaches were used: (1) sequence analyses to predict where the proteins localize in the cell, and (2) identifying orthologs in other species using BLAST. To determine the potential localization sites of proteins encoded by *comNOPQ* genes, we used the sequence analysis algorithms TMHMM v2.0 (31, 55) and SignalP (8, 37, 38) to predict topological features such as transmembrane helices and signal peptides. PSORT-B (23) was used to predict protein location sites for the five Gram-negative cellular compartments (cytoplasm, inner-membrane, periplasm, outer-membrane, or outside the cell). The final PSORT-B localization predictions were inconclusive for each of the four ComNOPQ proteins. Thus, a combination of the transmembrane helix and signal peptide prediction algorithms were used to determine possible protein localization sites.

Predictions from SignalP version 3.0 (8) indicated that ComN does not contain a signal peptide, while ComO has a non-cleavable signal anchor rather than a cleavable signal peptide. The TMHMM algorithm predicted that both ComN and ComO contain single transmembrane helices. Therefore, ComN and ComO appear to be inner-membrane proteins. TMHMM also predicts that ComQ and ComP have single transmembrane helices near the N-terminal regions, however, SignalP indicates that these hydrophobic segments are parts of cleavable signal peptides. As such, ComQ and ComP may localize to either the periplasm, the outer-membrane, or outside the cell.

Functional predictions of ComNOPQ proteins by ortholog comparisons

Our expectation is that *H. influenzae* genes involved in competence will be conserved in related species with similar competence mechanisms. Using BLAST (3) and the TIGR database (*www.tigr.org*), genes with strong sequence similarities to *comNOPQ* genes were found within the Pasteurellaceae family. In addition, *comO* (HI0939) was similar to sequences (*ppdB*, or prepilin peptidase dependent protein B) in *Yersinia pestis* and *Photorhabdus luminescens*. This led us to use PSI-BLAST (4, 5), a program that is more sensitive than BLAST in detecting weak sequence similarities by using iterated searches for orthologs of *comNOPQ* genes in other species.

With PSI-BLAST, weak orthologs of *comN* and *comO* were found in many bacterial species. These were primarily in the Proteobacteria (beta-, delta- and gamma-subdivisions), but also in the Cyanobacteria (*Nostoc* sp.) and Firmicutes (*Bacillus* spp., and *Streptococcus pnuemoniae*). In *E. coli* and all other sequenced species in the Enterobacteriaceae, *comN* and *comO* orthologs are named *ppdA* and *ppdB*, respectively. These genes encode proteins belonging to the PulG and PulJ families, pilin-like components of the type II secretion system, which are required for pullulanase secretion in *Klebsiella pneumoniae* and pectate lyase secretion in *Erwinia chrysanthemi* (28, 42, 50, 59). Interestingly, *comN* has weak similarity to *comGD* and *cglC*, which encode pilin-like proteins required for DNA uptake in *Bacillus* spp. and *Streptococcus pneumoniae*, respectively (12, 13, 39). Alignments of ComN with ComGD proteins from *Bacillus* spp. showed patchy similarity along the lengths of the proteins, with the highest similarity near the N-terminal region. In particular, ComN has a putative PilD peptidase

cleavage site near the N-terminus that is similar to the sequence in the *Bacillus* ComGD and pilins from other species (Figure-4). The peptidase cleaves the prepilin after the glycine residue in the –1 position, and the residue in the +1 position of mature pilins is either a phenyalanine or methionine that can be N-methylated by the PilD peptidase (43, 56). The N-methylation may be needed for pilus assembly, but does not seem to be essential for substrate recognition by the prepilin peptidase (43). ComO also has a potential PilD peptidase cleavage site (KGQTL). The type IV prepilin signal peptide is similar in amino acid composition as the classical signal peptide for type II secretion, except that the prepilin signal peptide is cleaved between the N-terminal and hydrophobic regions (43). Thus, the localization predictions made by SignalP and TMHMM did not account for alternative mechanisms used by the general secretion pathway.

PSI-BLAST did not detect any sequences significantly similar to *comP* (HI0940) or *comQ* (HI0941) outside the Pasteurellaceae. In the Enterobacteriaceae, *ppdA* and *ppdB* (*comN* and *comO* orthologs) are present in a four-gene cluster with *ygdB* and *ppdC*. BLAST searches directly against the *E. coli* K-12 genome did not detect YgdB or PpdC as orthologs to ComP or ComQ, and alignments using CLUSTALW indicated that ComP and ComQ are probably not related to YgdB and PpdC.

The localization predictions and ortholog comparisons suggest that the ComNOPQ proteins could be components of the DNA binding and uptake machinery or function to assemble such apparatus. The construction of non-polar mutants of the *comNOPQ* genes would be needed to test these hypotheses.

```
                                  -1 +1     ****************
Hi ComN                        MQKG  MTLVELLIGLAIISIALNFAVPLWKTDSPKTILAKEQHRLYLFL 48
Pm ComN                        MKRG  FTLLEILMVLLLISSMLLVVLPNWTRVIDFISFEQEQRQLWIFL 48
Aa PilA                MKIQSAFISLKFVKKG  FTLIELMIVIAIVAILATVAVPSYQNYTKKAAVSELLQASAPLR 60
Ec PpdA                      MKTQRG  YTLIETLVAMLILVMLSASGLYGWQYWQQSQRLWQTASQARDYL 50
Bs ComGD                   MNIKLNEEKG  FTLLESLLVLSLASILLVAVFTTLPPAYDNTAVRQAASQLKNDI 54
Sp CglC                 MKKMMTFLKKAKVKA  FTLVEMLVVLLIISVLFLLFVPNLTK--QKEAVNDKGKAAVVKV 57
Ng PilE                      MNTLQKG  FTLIELMIVIAIVGILAAAALPAYQDYTARAQVSEAILLAEGQK 51
Hi PilA                    MKLTTLQTLKKG  FTLIELMIVIAIIAILATIAIPSYQNYTKKAAVSELLQASAPYK 56
```

Figure 3.4: ComN contains a putative prepilin peptidase cleavage site.

Multiple-sequence alignments (57) of ComN to weak orthologues revealed similarity near the N-terminal ends. The potential peptidase cleavage sites (KGFTL) are separated by gaps in the protein sequences. The peptidase cleaves the prepilin after the glycine in the –1 position. The +1 position indicates the residue (typically phenylalanine or methionine) in processed pilins that is potentially N-methylated by the prepilin peptidase (56). The asterisks indicate the hydrophobic domains that follow the processing site. All sequences are from the TIGR or GenBank databases. Pilin-like proteins are: *H. influenzae* (Hi) HI0938; *P. multocida* (Pm) PM0965; *Actinobacillus actinomycetemcomitans* (Aa) PilA; *E. coli* K-12 (Ec) b2826; *B. subtilis* (Bs) Bsu2468; *S. pneumoniae* (Sp) CglC. Pilins from *N. gonhorreae* (Ng PilE) and *H. influenzae* (Hi PilA) are also included.

3.5 DISCUSSION

We have presented evidence that a previously unknown competence operon, *comNOPQ*, contains genes that are essential for competence in *H. influenzae*. It is interesting that these genes are related to prepilins, or type-IV pilus biogenesis proteins, since pilin-like proteins are involved in the competence mechanisms of *H. influenzae* and other bacteria (1, 9, 14, 18, 26, 48). However, *H. influenzae* has not been reported to have pili and the precise roles of these pilin-like components during the uptake process remain to be determined.

We have not ruled out the possibility that the *comN* deletion has polar effects on downstream genes in the *comNOPQ* operon. Thus, we can only conclude that one or more genes in the *comNOPQ* operon is involved in competence. The results from the transformation and DNA uptake assays demonstrate that the *comNOPQ* operon contains genes with essential functions for competence in *H. influenzae*. The competence phenotypes observed indicate that these genes may either be components of the DNA binding and uptake machinery, or function in the assembly of the DNA receptor and uptake machinery.

The *comN* and *comO* genes are related to components in the Pul secreton, PulG and PulJ, respectively. Both PulG and PulJ are required for pullulanase secretion in *Klebsiella pneumoniae* (28). Recently, it was shown that PulG, an abundant pseudo-pilin, is able to form a periplasmic pilus-like structure (50, 59). The function of the PulG pilus remains unknown, but is thought to be responsible for generating the force necessary to transport

pullulanase through secretin channels (59). In this case, ComN may have a similar function during competence in *H. influenzae*, but for transporting the DNA receptor to the cell surface. At present, little is known about PulJ, a minor pseudopilin, except that it is not needed for PulG pilus-assembly, but is necessary for pullulanase secretion (59). Therefore, it is uncertain what function ComO has during competence. Their relatedness to PulG and PulJ proteins suggest that ComN and ComO could assemble into pili-like structures in competent *H. influenzae*.

The most interesting feature of natural competence in *H. influenzae* is that this organism takes up DNA preferentially from its own species (51). This mechanism requires the recognition of uptake signal sequences (USSs) on exogenous DNA by sequence specific receptors (15, 51, 52). However, despite extensive research on uptake specificity in *H. influenzae* (6, 15, 25, 53), the identities of sequence-specific DNA receptors in the Pasteurellaceae remain unknown. Previously, candidates for the USS-specific DNA receptor (Chapter 2) were identified. Here, we show that ComN and ComO contain potential prepilin peptidase cleavage sites, indicating that these proteins might be transported to the cell surface as components of the DNA binding and uptake machinery. It is possible that the USS-receptor and DNA uptake apparatus is a pilin-like structure that traverses outermembrane in *H. influenzae*. We have not ruled out whether the *comN* deletion has polar effects on downstream genes in the *comNOPQ* operon. It will be interesting to determine if each of the *comNOPQ* genes function during competence, and whether they can affect uptake specificity in *H. influenzae*.

3.6 REFERENCES:

1. **Aas, F. E., M. Wolfgang, S. Frye, S. Dunham, C. Lovold, and M. Koomey.** 2002. Competence for natural transformation in *Neisseria gonorrhoeae*: components of DNA binding and uptake linked to type IV pilus expression. Mol Microbiol **46:**749-60.
2. **Alexander, H. E., and G. Leidy.** 1951. Determination of inherited traits of H. influenzae by desoxyribonucleic acid fractions isolated from type-specific cells. J Exp Med **93:**345-59.
3. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J Mol Biol **215:**403-10.
4. **Altschul, S. F., and E. V. Koonin.** 1998. Iterated profile searches with PSI-BLAST--a tool for discovery in protein databases. Trends Biochem Sci **23:**444-7.
5. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25:**3389-402.
6. **Bakkali, M., T. Y. Chen, H. C. Lee, and R. J. Redfield.** 2004. Evolutionary stability of DNA uptake signal sequences in the Pasteurellaceae. Proc Natl Acad Sci U S A **101:**4513-8.
7. **Barcak, G. J., M. S. Chandler, R. J. Redfield, and J. F. Tomb.** 1991. Genetic systems in *Haemophilus influenzae*. Methods Enzymol **204:**321-42.
8. **Bendtsen, J. D., H. Nielsen, G. von Heijne, and S. Brunak.** 2004. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol **340:**783-95.
9. **Busch, S., C. Rosenplanter, and B. Averhoff.** 1999. Identification and characterization of ComE and ComF, two novel pilin-like competence factors involved in natural transformation of Acinetobacter sp. strain BD413. Appl Environ Microbiol **65:**4568-74.
10. **Catlin, B. W., J. W. Bendler, 3rd, and S. H. Goodgal.** 1972. The type b capsulation locus of *Haemophilus influenzae*: map location and size. J Gen Microbiol **70:**411-22.
11. **Chen, W. J., L. Gross, K. E. Joho, and W. T. McAllister.** 1992. A modified kanamycin-resistance cassette to facilitate two-codon insertion mutagenesis. Gene **111:**143-4.
12. **Chung, Y. S., F. Breidt, and D. Dubnau.** 1998. Cell surface localization and processing of the ComG proteins, required for DNA binding during transformation of Bacillus subtilis. Mol Microbiol **29:**905-13.
13. **Chung, Y. S., and D. Dubnau.** 1998. All seven comG open reading frames are required for DNA binding during transformation of competent Bacillus subtilis. J Bacteriol **180:**41-5.
14. **Collins, R. F., L. Davidsen, J. P. Derrick, R. C. Ford, and T. Tonjum.** 2001. Analysis of the PilQ secretin from Neisseria meningitidis by transmission electron microscopy reveals a dodecameric quaternary structure. J Bacteriol **183:**3825-32.
15. **Deich, R. A., and H. O. Smith.** 1980. Mechanism of homospecific DNA uptake in *Haemophilus influenzae* transformation. Mol Gen Genet **177:**369-74.

16. **Dougherty, B. A., and H. O. Smith.** 1999. Identification of *Haemophilus influenzae* Rd transformation genes using cassette mutagenesis. Microbiology **145 (Pt 2):**401-9.
17. **Drake, S. L., and M. Koomey.** 1995. The product of the *pilQ* gene is essential for the biogenesis of type IV pili in *Neisseria gonorrhoeae*. Mol Microbiol **18:**975-86.
18. **Dubnau, D.** 1997. Binding and transport of transforming DNA by *Bacillus subtilis*: the role of type-IV pilin-like proteins--a review. Gene **192:**191-8.
19. **Dubnau, D.** 1999. DNA uptake in bacteria. Annu Rev Microbiol **53:**217-44.
20. **Engels, W. R.** 1993. Contributing software to the internet: the Amplify program. Trends Biochem Sci **18:**448-50.
21. **Finkel, S. E., and R. Kolter.** 2001. DNA as a nutrient: novel role for bacterial competence gene homologs. J Bacteriol **183:**6288-93.
22. **Friedrich, A., T. Hartsch, and B. Averhoff.** 2001. Natural transformation in mesophilic and thermophilic bacteria: identification and characterization of novel, closely related competence genes in Acinetobacter sp. strain BD413 and Thermus thermophilus HB27. Appl Environ Microbiol **67:**3140-8.
23. **Gardy, J. L., C. Spencer, K. Wang, M. Ester, G. E. Tusnady, I. Simon, S. Hua, K. deFays, C. Lambert, K. Nakai, and F. S. Brinkman.** 2003. PSORT-B: Improving protein subcellular localization prediction for Gram-negative bacteria. Nucleic Acids Res **31:**3613-7.
24. **Goodgal, S. H., and R. M. Herriott.** 1961. Studies on transformations of Hemophilus influenzae. I. Competence. J Gen Physiol **44:**1201-27.
25. **Goodgal, S. H., and M. A. Mitchell.** 1990. Sequence and uptake specificity of cloned sonicated fragments of *Haemophilus influenzae* DNA. J Bacteriol **172:**5924-8.
26. **Graupner, S., V. Frey, R. Hashemi, M. G. Lorenz, G. Brandes, and W. Wackernagel.** 2000. Type IV pilus genes pilA and pilC of Pseudomonas stutzeri are required for natural genetic transformation, and pilA can be replaced by corresponding genes from nontransformable species. J Bacteriol **182:**2184-90.
27. **Hanahan, D.** 1983. Studies on transformation of Escherichia coli with plasmids. J Mol Biol **166:**557-80.
28. **He, S. Y., M. Lindeberg, A. K. Chatterjee, and A. Collmer.** 1991. Cloned Erwinia chrysanthemi out genes enable Escherichia coli to selectively secrete a diverse family of heterologous proteins to its milieu. Proc Natl Acad Sci U S A **88:**1079-83.
29. **Herriott, R. M., E. M. Meyer, and M. Vogt.** 1970. Defined nongrowth media for stage II development of competence in *Haemophilus influenzae*. J Bacteriol **101:**517-24.
30. **Karudapuram, S., and G. J. Barcak.** 1997. The *Haemophilus influenzae* dprABC genes constitute a competence-inducible operon that requires the product of the tfoX (sxy) gene for transcriptional activation. J Bacteriol **179:**4815-20.
31. **Krogh, A., B. Larsson, G. von Heijne, and E. L. Sonnhammer.** 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol **305:**567-80.

32. **Link, C., S. Eickernjager, D. Porstendorfer, and B. Averhoff.** 1998. Identification and characterization of a novel competence gene, comC, required for DNA binding and uptake in Acinetobacter sp. strain BD413. J Bacteriol **180:**1592-5.
33. **Lorenz, M. G., and W. Wackernagel.** 1994. Bacterial gene transfer by natural genetic transformation in the environment. Microbiol Rev **58:**563-602.
34. **Macfadyen, L. P.** 2000. Regulation of competence development in *Haemophilus influenzae*. J Theor Biol **207:**349-59.
35. **Marck, C.** 1988. 'DNA Strider': a 'C' program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. Nucleic Acids Res **16:**1829-36.
36. **Mathis, L. S., and J. J. Scocca.** 1982. *Haemophilus influenzae* and *Neisseria gonorrhoeae* recognize different specificity determinants in the DNA uptake step of genetic transformation. J Gen Microbiol **128 (Pt 5):**1159-61.
37. **Nielsen, H., S. Brunak, and G. von Heijne.** 1999. Machine learning approaches for the prediction of signal peptides and other protein sorting signals. Protein Eng **12:**3-9.
38. **Nielsen, H., J. Engelbrecht, S. Brunak, and G. von Heijne.** 1997. A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Int J Neural Syst **8:**581-99.
39. **Pestova, E. V., and D. A. Morrison.** 1998. Isolation and characterization of three Streptococcus pneumoniae transformation-specific loci by use of a lacZ reporter insertion vector. J Bacteriol **180:**2701-10.
40. **Poje, G., and R. J. Redfield.** 2003. General methods for culturing *Haemophilus influenzae*. Methods Mol Med **71:**51-6.
41. **Poje, G., and R. J. Redfield.** 2003. Transformation of *Haemophilus influenzae*. Methods Mol Med **71:**57-70.
42. **Possot, O. M., G. Vignon, N. Bomchil, F. Ebel, and A. P. Pugsley.** 2000. Multiple interactions between pullulanase secreton components involved in stabilization and cytoplasmic membrane association of PulE. J Bacteriol **182:**2142-52.
43. **Pugsley, A. P.** 1993. The complete general secretory pathway in gram-negative bacteria. Microbiol Rev **57:**50-108.
44. **Redfield, R. J.** 1988. Evolution of bacterial transformation: is sex with dead cells ever better than no sex at all? Genetics **119:**213-21.
45. **Redfield, R. J.** 1993. Evolution of natural transformation: testing the DNA repair hypothesis in *Bacillus subtilis* and *Haemophilus influenzae*. Genetics **133:**755-61.
46. **Redfield, R. J.** 1993. Genes for breakfast: the have-your-cake-and-eat-it-too of bacterial transformation. J Hered **84:**400-4.
47. **Redfield, R. J., M. R. Schrag, and A. M. Dean.** 1997. The evolution of bacterial transformation: sex with poor relations. Genetics **146:**27-38.
48. **Rudel, T., D. Facius, R. Barten, I. Scheuerpflug, E. Nonnenmacher, and T. F. Meyer.** 1995. Role of pili and the phase-variable PilC protein in natural competence for transformation of *Neisseria gonorrhoeae*. Proc Natl Acad Sci U S A **92:**7986-90.

49. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular Cloning, A Laboratory Manual, 2nd Edt. Cold Spring Harbour Laboratory Press, USA.
50. **Sauvonnet, N., G. Vignon, A. P. Pugsley, and P. Gounon.** 2000. Pilus formation and protein secretion by the same machinery in Escherichia coli. Embo J **19:**2221-8.
51. **Scocca, J. J., R. L. Poland, and K. C. Zoon.** 1974. Specificity in deoxyribonucleic acid uptake by transformable Haemophilus influenzae. J Bacteriol **118:**369-73.
52. **Sisco, K. L., and H. O. Smith.** 1979. Sequence-specific DNA uptake in *Haemophilus* transformation. Proc Natl Acad Sci U S A **76:**972-6.
53. **Smith, H. O., J. F. Tomb, B. A. Dougherty, R. D. Fleischmann, and J. C. Venter.** 1995. Frequency and distribution of DNA uptake signal sequences in the *Haemophilus influenzae* Rd genome. Science **269:**538-40.
54. **Solomon, J. M., and A. D. Grossman.** 1996. Who's competent and when: regulation of natural genetic competence in bacteria. Trends Genet **12:**150-5.
55. **Sonnhammer, E. L., G. von Heijne, and A. Krogh.** 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. Proc Int Conf Intell Syst Mol Biol **6:**175-82.
56. **Strom, M. S., D. N. Nunn, and S. Lory.** 1993. A single bifunctional enzyme, PilD, catalyzes cleavage and N-methylation of proteins belonging to the type IV pilin family. Proc Natl Acad Sci U S A **90:**2404-8.
57. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res **22:**4673-80.
58. **Tomb, J. F., H. el-Hajj, and H. O. Smith.** 1991. Nucleotide sequence of a cluster of genes involved in the transformation of *Haemophilus influenzae* Rd. Gene **104:**1-10.
59. **Vignon, G., R. Kohler, E. Larquet, S. Giroux, M. C. Prevost, P. Roux, and A. P. Pugsley.** 2003. Type IV-like pili formed by the type II secreton: specificity, composition, bundling, polar localization, and surface presentation of peptides. J Bacteriol **185:**3416-28.
60. **Wang, Y., W. Shi, W. Chen, and C. Chen.** 2003. Type IV pilus gene homologs pilABCD are required for natural transformation in *Actinobacillus actinomycetemcomitans*. Gene **312:**249-55.
61. **Wilcox, K. W., and H. O. Smith.** 1975. Isolation and characterization of mutants of *Haemophilus influenzae* deficient in an adenosine 5'-triphosphate-dependent deoxyribonuclease activity. J Bacteriol **122:**443-53.
62. **Williams, P., W. L. Hung, and R. J. Redfield.** 1996. Cell transfer by filtration: evaluation of protocols for transformational competence. FEMS Microbiol Lett **137:**183-7.
63. **Wolfgang, M., P. Lauer, H. S. Park, L. Brossay, J. Hebert, and M. Koomey.** 1998. PilT mutations lead to simultaneous defects in competence for natural transformation and twitching motility in piliated *Neisseria gonorrhoeae*. Mol Microbiol **29:**321-30.

Chapter 4: Functional Predictions of CRE Regulated Genes in *Haemophilus influenzae.*

4.1 Theoretical protein localization of *H. influenzae* competence genes

Determining where competence proteins localize in the cell (i.e., the cytoplasm, inner membrane, periplasm, outer membrane, or secreted outside the cell) may assist in understanding what their functions are in competence. Protein localization predictions were made using sequence analysis programs SignalP (signal peptide prediction) (14, 15), TMHMM (transmembrane helix prediction) (12), and PSORT-B (predicts protein localizations to one or more of the five possible compartments in Gram-negative bacteria) (8). The raw data (probabilities, topological features, and possible localization sites) generated from each of these programs used for all CRE regulated genes in *H. influenzae* are provided in Appendix B. A summary of the analyses from each algorithm and a final overall prediction are given in Table 4.1. Protein localization predictions for some of the competence proteins are discussed below. I suggest possible roles for each competence protein with the goal of identifying those that are likely to be part of the DNA binding, uptake, and translocation machinery, as well as candidates responsible for specifically recognizing the USS.

4.2 Identification of candidates for the USS-specific DNA receptor

The most likely candidates for the USS-specific receptor in *H. influenzae* are CRE regulated competence genes whose proteins are localized to the cell surface. The presence of a signal peptide detected by SignalP and a final localization to the outer

membrane or "outside the cell" predicted by PSORT-B are the desired predictions for a USS-specific DNA receptor. Unfortunately, most of the final predictions by PSORT-B for the CRE regulated genes were inconclusive (Table 4.1). As noted by Gardy et al. (8), PSORT-B is less effective in predicting the localization of periplasmic proteins, proteins containing less than three transmembrane helices, and proteins that have multiple localization sites.

Several of the CRE genes are predicted by TMHMM v.2.0 to encode proteins with only single transmembrane helices (Table 4.1). These are *comNOPQ*, *comB*, *comC*, and *pilA*. However, transmembrane helices (TMHs) and signal peptides (SPs) may be predicted for the same amino acid region in the same protein (12). This is due to the similarity in amino acid composition of the hydrophobic region in TMHs and SPs. Signal peptides that are not cleaved are known as "signal anchors" and proteins with signal anchors remain embedded in the inner-membrane. Whether or not a signal peptide is likely to be cleaved or not cleaved can be predicted by SignalP version 2.0. If the S-score predicts a signal peptide (the S-score is higher than the 0.44 cutoff value), while the Y-score does not (the Y-score is less than the 0.32 cutoff value), then the protein is likely to be a signal anchor (14, 15). Therefore, closer inspection of the probabilities generated by SignalP may assist in determining the most likely localization sites for competence proteins.

As predicted by SignalP version 2.0 (see Appendix B), the following CRE regulated genes encode proteins with signal peptides: *comNOPQ*, *comB*, *comD*, *comE*, *comE1*, *pilA*, and HI1182. However, comparing the Y-scores and S-scores to threshold values,

the proteins encoded by *comB*, *pilA*, *comN*, and *comO* are likely to have signal anchors (i.e., these proteins are expected to localize to the inner membrane only). Competence proteins were re-analyzed using the recently released SignalP version 3.0 (1), which increases the accuracy of cleavage site predictions and reduces false positives. The final predictions for all competence proteins were similar using SignalP v.3.0 as compared to version 2.0, except for ComN in which a signal peptide was not detected by version 3.0. Overall, the SignalP results are consistent with the data generated from TMHMM, which together suggest that ComN, ComO, ComB, and PilA are inner membrane proteins.

The SignalP algorithm is only capable of identifying classical signal peptides that are cleaved by signal peptidase I (SPaseI) (1). The classic signal peptide contains an N-terminal region, a hydrophobic core, and a C-terminal region followed by an SPaseI cleavage site (5, 19). There are two other types of SPs that are similar to the classical SP type: lipoprotein SPs and type IV prepilin SPs (19). Lipoproteins have shorter signal peptides that are cleaved by signal peptidase II (SPaseII) (19). To determine if there were any SPaseII cleavage sites in *H. influenzae* competence proteins, I used LipoP (10) and found that only ComD has a potential SPaseII cleavage site. Type IV prepilin SPs are similar to the classical SPs, but are cleaved by prepilin peptidase between the N-terminus and the hydrophobic core (17, 19, 20). Thus, SignalP may detect the presence of SPs in prepilins or prepilin-like proteins, but will give an incorrect cleavage site for pilins and pilin-like proteins. Since the N-terminal ends of unprocessed PilA, ComN, and ComO contain potential pilin peptidase cleavage sites (Chapter 3), it seems likely that these proteins are not inner-membrane proteins, as predicted by SignalP or TMHMM, but are

transported to the periplasm, the outer membrane, or outside the cell. Therefore, PilA, ComN, and ComO are included as USS-receptor candidates.

In Gram-negative bacteria, secreted proteins may be transported by one of several terminal branches in the general secretory pathway (GSP) (19), including the type II secretion system that uses signal peptides for targeting proteins to the inner membrane or to other compartments. Since non-classical signal peptides or leader-less secreted proteins are not identified by SignalP, a negative classification does not necessarily mean a protein is a non-secreted protein (1). Thus, it is possible that some of the competence proteins identified as being cytoplasmic (Table 4.1) could be secreted. At present, there are no prediction methods to make this determination; therefore the simplest explanation is that these proteins are cytoplasmic.

The remaining CRE regulated genes that encode proteins with signal peptides are *comD*, *comE*, *comE1*, *comP*, and *comQ* (Table 4.1). These genes, and the pilin-like genes *pilA*, *comN*, and *comO*, are the most likely candidates for encoding the USS-specific receptor. In Chapter 2, I ruled out *comE1* (HI1008) as a USS-receptor candidate, since the *comE1* knockout is capable of uptake specificity during transformation. I would also rule out *comP* as a USS-receptor gene, since orthologs were not found in *Actinobacillus actinomycetemcomitans* – a species known to have the USS-specific uptake mechanism.

The *comE* (HI0435) gene is part of the *comABCDEF* operon (11, 13) and a transposon insertion in *comE* eliminates DNA binding, uptake, and transformation (21). The *comE*

knockout does not appear to have polar effects on *comF*, since the disruption of *comF* only eliminates transformation and not DNA binding and uptake (21). BLAST searches revealed the presence of ComE orthologs, such as PilQ, in many bacterial species (7, 22). The ComE and PilQ proteins belong to the PulD superfamily known as "secretins" (3, 4, 16). PilQ forms a dodecameric ring and is involved in the assembly of the type II secretion (type IV pili) apparatus in neisserial species (3, 4). Interestingly, PilQ is essential for competence in *Neisseria gonorrhoeae* (7). It is likely that ComE plays a similar role in *H. influenzae*. For example, ComE could function in the assembly of the DNA binding and uptake apparatus during competence development. It is also possible that ComE binds directly to incoming DNA and acts as a channel for DNA transport across the outer membrane and into the periplasm. However, the precise function of ComE in competence remains to be experimentally determined. Currently, the role of ComE as the USS-receptor is being tested in the Redfield lab.

At present, mutants of the USS-receptor candidates *comD*, *comO*, and *comQ* do not exist. Although these candidates are part of CRE regulated operons, it is not known whether these genes directly function in competence. It is possible that disruption of *comN* (Chapter 3) has polar effects on the other genes in operon *comNOPQ*. In this case, *comN* might not be a part of the competence mechanism. One way to test if *comN* is required for competence is to complement the deleted *comN* with a plasmid that carries and expresses the wild type *comN* gene. This approach could be used for the other competence genes and for USS-receptor candidates for which mutants are not yet available.

Table 4.1: Localization Predictions for Competence Proteins in *H. influenzae*.*

Gene	TIGR HI#	SignalP v2.0	TMHMM v2.0 (No. of TMH)	PSORT-B**	Overall Prediction
comN	0938	SA	TMH (1)	Unknown (C)	PS/OM/SEC
comO	0939	SA	TMH (1)	Unknown (TMH)	PS/OM/SEC
comP	0940	SP	TMH (1)	Unknown (TMH)	PS/OM/SEC
comQ	0941	SP	TMH (1)	Unknown (TMH)	PS/OM/SEC
comA	0439	None	None	Unknown (C)	C
comB	0438	SA	TMH (1)	Unknown (TMH)	IM
comC	0437	None	TMH (1)	Unknown (TMH)	IM
comD	0436	SP	None	Unknown (C)	PS/OM/SEC
comE	0435	SP	TMH (1)	OM	OM
comF	0434	None	None	Unknown (Unk.)	C
comE1	1008	SP	None	Unknown (SP/C)	PS/OM/SEC
comM	1117	None	None	Unknown (Unk.)	C
dprA	0985	None	None	Unknown (Unk.)	C
pilA	0299	SA	TMH (1)	Unknown (Unk.)	PS/OM/SEC
pilB	0298	None	None	C	C
pilC	0297	None	TMH (3)	IM	IM
pilD	0296	None	TMH (6)	IM	IM
radC	0952	None	None	Unknown (C)	C
rec2	0061	None	TMH (11)	IM	IM
ssb	0250	None	None	Unknown (Unk.)	C
Hyp.	0365	None	None	Unknown (C)	C
Hyp.	0659	None	None	Unknown (Unk.)	C
Hyp.	0660	None	None	Unknown (C)	C
Hyp.	1162	None	None	C	C
Hyp.	1182	SP	None	Unknown (SP/C)	PS/OM/SEC
Hyp.	1183	None	None	Unknown (Unk.)	C

*SP, signal peptide; SA, signal anchor; TMH, transmembrane helix; C, cytoplasmic; Unk, unknown; IM, inner membrane; PS, periplasm; OM, outer membrane; SEC, secreted

**Abbreviations in parentheses are possible localization sites or features predicted by PSORT-B.

4.3 CHAPTER FOUR REFERENCES:

1. **Bendtsen, J. D., H. Nielsen, G. von Heijne, and S. Brunak.** 2004. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol **340:**783-95.
2. **Caster, J. H., E. H. Postel, and S. H. Goodgal.** 1970. Competence mutants: isolation of transformation deficient strains of *Haemophilus influenzae*. Nature **227:**515-7.
3. **Collins, R. F., L. Davidsen, J. P. Derrick, R. C. Ford, and T. Tonjum.** 2001. Analysis of the PilQ secretin from Neisseria meningitidis by transmission electron microscopy reveals a dodecameric quaternary structure. J Bacteriol **183:**3825-32.
4. **Collins, R. F., R. C. Ford, A. Kitmitto, R. O. Olsen, T. Tonjum, and J. P. Derrick.** 2003. Three-dimensional structure of the Neisseria meningitidis secretin PilQ determined from negative-stain transmission electron microscopy. J Bacteriol **185:**2611-7.
5. **Dev, I. K., P. H. Ray, and P. Novak.** 1990. Minimum substrate sequence for signal peptidase I of Escherichia coli. J Biol Chem **265:**20069-72.
6. **Dougherty, B. A., and H. O. Smith.** 1999. Identification of *Haemophilus influenzae* Rd transformation genes using cassette mutagenesis. Microbiology **145 (Pt 2):**401-9.
7. **Drake, S. L., and M. Koomey.** 1995. The product of the *pilQ* gene is essential for the biogenesis of type IV pili in *Neisseria gonorrhoeae*. Mol Microbiol **18:**975-86.
8. **Gardy, J. L., C. Spencer, K. Wang, M. Ester, G. E. Tusnady, I. Simon, S. Hua, K. deFays, C. Lambert, K. Nakai, and F. S. Brinkman.** 2003. PSORT-B: Improving protein subcellular localization prediction for Gram-negative bacteria. Nucleic Acids Res **31:**3613-7.
9. **Inamine, G. S., and D. Dubnau.** 1995. ComEA, a *Bacillus subtilis* integral membrane protein required for genetic transformation, is needed for both DNA binding and transport. J Bacteriol **177:**3045-51.
10. **Juncker, A. S., H. Willenbrock, G. Von Heijne, S. Brunak, H. Nielsen, and A. Krogh.** 2003. Prediction of lipoprotein signal peptides in Gram-negative bacteria. Protein Sci **12:**1652-62.
11. **Karudapuram, S., and G. J. Barcak.** 1997. The *Haemophilus influenzae* dprABC genes constitute a competence-inducible operon that requires the product of the tfoX (sxy) gene for transcriptional activation. J Bacteriol **179:**4815-20.
12. **Krogh, A., B. Larsson, G. von Heijne, and E. L. Sonnhammer.** 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol **305:**567-80.
13. **Macfadyen, L. P.** 2000. Regulation of competence development in *Haemophilus influenzae*. J Theor Biol **207:**349-59.
14. **Nielsen, H., S. Brunak, and G. von Heijne.** 1999. Machine learning approaches for the prediction of signal peptides and other protein sorting signals. Protein Eng **12:**3-9.
15. **Nielsen, H., J. Engelbrecht, S. Brunak, and G. von Heijne.** 1997. A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Int J Neural Syst **8:**581-99.

16. **Nouwen, N., H. Stahlberg, A. P. Pugsley, and A. Engel.** 2000. Domain structure of secretin PulD revealed by limited proteolysis and electron microscopy. Embo J **19:**2229-36.
17. **Nunn, D. N., and S. Lory.** 1991. Product of the Pseudomonas aeruginosa gene pilD is a prepilin leader peptidase. Proc Natl Acad Sci U S A **88:**3281-5.
18. **Provvedi, R., and D. Dubnau.** 1999. ComEA is a DNA receptor for transformation of competent *Bacillus subtilis*. Mol Microbiol **31:**271-80.
19. **Pugsley, A. P.** 1993. The complete general secretory pathway in gram-negative bacteria. Microbiol Rev **57:**50-108.
20. **Strom, M. S., and S. Lory.** 1992. Kinetics and sequence specificity of processing of prepilin by PilD, the type IV leader peptidase of Pseudomonas aeruginosa. J Bacteriol **174:**7345-51.
21. **Tomb, J. F., G. J. Barcak, M. S. Chandler, R. J. Redfield, and H. O. Smith.** 1989. Transposon mutagenesis, characterization, and cloning of transformation genes of *Haemophilus influenzae* Rd. J Bacteriol **171:**3796-802.
22. **Tomb, J. F., H. el-Hajj, and H. O. Smith.** 1991. Nucleotide sequence of a cluster of genes involved in the transformation of *Haemophilus influenzae* Rd. Gene **104:**1-10.

APPENDIX A: Plasmid Construction and HI0938 Knockout Procedures.

This appendix section includes the overall cloning strategy (Figure-A.1), details of the plasmids constructed (Figures-A.2 to –A.20), and of procedures for the chromosomal HI0938 knockout strains generated (Figures-A.21 to –A.29). In this study, LEFT and RIGHT refer to the 1 kb upstream and downstream flanking DNA regions from the HI0938 start and stop codons, respectively. KAN refers to the kanamycin resistance cassette from pWJC3.

A.1 Construction of plasmids pLEFT, pKAN, and pRIGHT

In steps 1, 2, and 3 (Figure-A.1), PCR was used to amplify the LEFT, RIGHT, and KAN DNA segments (Figure-A.2). The PCR products were gel purified and ligated into the appropriate vectors (i.e. LEFT and RIGHT each into pGEM-T Easy, and KAN into pTOPO) to make the best use of the multiple cloning sites, and then cloned into *E. coli*. Transformants were selected for using ampicillin LB plates, and the pLEFT, pRIGHT, and pKAN clones were chosen for colony PCR analysis based on Blue/White screening. Colony PCR was then used to detect and confirm the presence of the desired constructs in the clones (Figures-A.3 and -A.4 for pLEFT; -A.7 for pKAN; and -A.10 for pRIGHT), using primers corresponding to the appropriate inserts (i.e. primers P1 and P2 for LEFT; K1 and K2 for KAN; and P3 and P4 for RIGHT). To determine insert orientations, restriction maps and digests for pLEFT (Figures-A.5 and -A.6), pKAN (-A.8 and –A.9), and pRIGHT (-A.11 and -A.12) are shown. The *E. coli* clones carrying the desired pLEFT, pRIGHT, and pKAN plasmids were given the strain numbers RR1160, RR1161, and RR1162, respectively.

The colony PCR results for pLEFT clones (RR1160) are shown in figures-A.3 and -A.4. Using P1 and P2 primers on purified DNA or on colonies, the expected band of 1124 bp for the LEFT fragment was observed for the colony tested (Figure-A.3, lanes 1, 2 and 3), as well as for the positive controls using gel purified LEFT DNA or a KW20 colony. No amplification was observed, as expected, when DH5α DNA was used as the template source in a negative control. Eight additional colonies from an independent transformation experiment were checked for the presence of the LEFT insert (Figure-A.4), of which three had tested positive (lanes 2, 4, and 6, Figure-A.4). To determine the orientation of the LEFT inserts, restriction digests using *PstI* was done on pLEFT plasmid DNA (Figures-A.5 and –A.6). The expected bands for LEFT in the correct orientation, are 1147 bp for the LEFT fragment and 2993 bp for the vector (Figure-A.5). For LEFT in the opposite orientation, 23 bp and 4117 bp bands are expected. Only one of the four pLEFT plasmids isolated had the LEFT insert in the correct orientation (RR1160), while the others had LEFT in the reverse orientation (Figure-A.6).

For putative pKAN transformants, the presence of the KAN insert was detected in all sixteen clones tested with colony PCR (Figure-A.7). Restriction digests using *BamHI* on purified pRIGHT DNA revealed that five of the clones (#K-9, K-11, K-13, K-14, & K-16) had the KAN insert in the correct orientation (Figures-A.8 and –A.9). One pKAN clone (K-11; RR1162) was used in the knockout procedure (Figure-A.9, lane 5).

For the putative pRIGHT clones, sixteen colonies were screened for the presence of the RIGHT insert with colony PCR using primers P3 and P4 (Figure-A.10). Five of the

clones showed relatively faint bands of the expected 997 bp size for the RIGHT amplicon, and were excluded from further testing (Figure-A.10). Minipreps were then done on nine of the remaining ten colonies, which had the RIGHT fragment, followed by restriction digests with *SspI* to determine the orientation of the RIGHT insert (Figures-A.11 and –A.12). Of the nine pRIGHT plasmids that were checked, two of them had the desired RIGHT insert orientation (pRIGHT-4 and pRIGHT-11). The pRIGHT-4 clone used to construct pLKR::ΔHI0938 was given the strain number RR1161.

A.2 Construction and cloning of the pLEFT_KAN and pLKR plasmids

Once the pLEFT, pKAN, and pRIGHT plasmids with the correct insert orientations were obtained, steps 4 and 5 (Figure-A.1) of the knockout procedure were implemented. pLEFT (from RR1160) and pKAN (RR1162) were cut with *BamHI*, and then gel purified. The KAN insert was then ligated into the pLEFT plasmid at the *BamHI* site and subsequently transformed into *E. coli* (Figure-A.1, step 4). Colony PCR was used to screen for kanamycin resistant colonies with the LEFT_KAN construct (Figures-A.13 and -A.14). The orientation of the KAN insert in pLEFT_KAN was determined with PCR using P1 and K2 primers (Figure-A.13), since no useful restriction sites were found (Figure-A.15). With primers P1 and K2, a band of 2100 bp is expected for pLEFT_KAN with KAN in the correct orientation, whereas no bands are expected when KAN is in the opposite orientation (Figure-A.13). Of eleven colonies tested, five where found to contain KAN in the desired orientation (Figure-A.14). One clone carrying the pLEFT_KAN plasmid was given the strain number RR1163.

The pLEFT_KAN (RR1163) and pRIGHT (RR1161) plasmids were cut with *PstI* (Figure-A.16), and the LEFT_KAN fragment was ligated into the *PstI* site of the pRIGHT vector. The ligation was then transformed directly into competent *E. coli*. Colony PCR was used to screen for clones with the pLEFT_KAN_RIGHT (pLKR) construct (Figures-A.17 and –A.18). In instances where colony PCR was suspected to be unreliable (e.g. no bands detected for LEFT_KAN in twenty colonies screened), restriction digests were done on plasmid DNA to determine insert orientations. However, in all cases where PCR had not detected LEFT_KAN in the correct orientation, restriction digests confirmed that LEFT_KAN was in the reverse orientation (data not shown). Of thirty-one clones tested, 25 of them contained the LEFT_KAN insert (Figure-A.18A). In parallel, colony PCR was used to determine the LEFT_KAN orientation in pLKR clones (Figure-A.18B). Only four clones (pLKR-45, -46 -48 and -49) were found to have the LEFT_KAN in the desired orientation, which was confirmed using restriction digests (Figures-A.19 and –A.20). The pLKR-45 and pLKR-46 clones were given strain numbers RR1164 and RR1165, respectively, and were used to generate HI0938 knockouts (step 6, Figure-A.1).

A.3 Generation of ΔHI0938 chromosomal mutants

A.3.1 Overview

Since HI0938 is part of a putative competence operon, the main goal of the present work was to construct *H. influenzae* strains with HI0938 chromosomal deletions, and test them for competence defects. Figures A.21 to A.29 are the results for the construction of HI0938 knockouts.

During the knockout procedures, several difficulties were encountered. PCR and colony PCR were frequently used during the screening of clones, and for confirming genotypes of plasmid constructs and chromosomal deletions. One problem was an inability to amplify fragments more than 2 kb from plasmid or chromosomal DNA templates using PCR. The reasons for this problem were unclear. Another problem was that colony PCR was unreliable. For example, 2kb segments could be amplified from plasmid DNA, but not from colonies. Thus, an alternative approach was used: primers (i.e. PC1 and PC2) that amplified DNA fragments of 1kb, instead of fragments greater than 2kb, were used to confirm plasmid and chromosomal genotypes.

One concern with constructing mutants is whether the target gene is an essential gene. We do not expect HI0938 to be an essential gene, as the microarray data suggest that HI0938 is effectively off under growth conditions and only turns on when competence is induced (data not shown). Furthermore, before construction of ΔHI0938 strains, we did not know what colony or growth phenotypes would be observed when HI0938 was deleted. In independent trials, kanamycin resistant colonies, carrying the putative HI0938

deletion, appeared as small (or "microcolonies") and normal sized colonies. The reasons for the colony sizes were unclear. Interestingly, clones that grew as microcolonies on *kan* selective plates grew to the typical wild type colony sizes when grown on plain plates.

During screens for putative ΔHI0938 clones, the microcolonies carried HI0938 deletions plus the KAN inserts, while normal sized putative ΔHI0938::*kan* colonies did not. Subsequently, using colony PCR to screen for HI0938 deletions and for the presence of the KAN insert was problematic when single microcolonies were tested. Thus, pools of microcolonies were screened with colony PCR for KAN inserts. Then, single clones grown in plain media were tested for HI0938 deletions. Finally, the ΔHI0938 strains obtained were tested for competence defects using transformation and DNA uptake assays (Chapter 4).

A.3.2 Construction of ΔHI0938 chromosomal mutants

Once the pLKR plasmid was obtained, it was used to generate strains carrying the chromosomal HI0938 deletion (Figure-A.1, step 6). To do this, the 3.3 kb LEFT_KAN_RIGHT (LKR) linear construct was cut out of pLKR::ΔHI0938 with *SacI* and *AatII* (Figure-A.24), and then transformed directly into competent wild type *H. influenzae*.

From the transformation step, kanamycin resistant colonies were screened for the KAN insert using colony PCR (Figure-A.22). Positive and negative controls for colony PCR gave results as expected (i.e. 997 bp band for KAN in positive controls, no bands present

in negative controls) (Figure-A.22). From seven pools of microcolonies tested, five were identified as having the KAN insert. In contrast, KAN was not detected in any of the large colonies that were screened (data not shown). Aliquots of overnight cultures of ΔHI0938::*kan* candidate pools were then frozen in glycerol at -80°C.

To prepare pure cultures, ten colonies from the one of the pools (containing 5 microcolonies) were chosen for screening, since this pool (as compared to the 30 microcolony pool) had the highest probability in which detection of the ΔHI0938::kan^R could be made. The ten colonies were checked for the presence of KAN with colony PCR using K1 and K2 primers (Figure-A.23). Positive controls included amplifying KAN from the pLKR plasmid and from a *purR* colony. The negative control used a wild type *H. influenzae* colony for the DNA template source and had no amplification in PCR. Of the ten colonies screened, only one had the KAN insert (Figure-A.23, lane 5). This clone was given the strain number RR1172.

To confirm the replacement of HI0938 with KAN in strain RR1172, colony PCR was attempted using different primer sets in parallel (i.e. P1/P2 for LEFT; K1/K2, KAN; P3/P4, RIGHT; P1/K2, LEFT_KAN; K1/P4, KAN_RIGHT, and P1/P4 for the full length LKR) (Figure-A.24). From these reactions, only the ~1 kb PCR products, representing the LEFT, KAN, and RIGHT fragments, were obtained (Figure-A.24, lanes 1, 2, 3), while none of the expected ~2kb for LEFT_KAN or KAN_RIGHT, or full length 3.3 kb LKR::ΔHI0938 fragments were amplified. Positive controls that used pLKR::ΔHI0938 as the template all worked (i.e. the 1 kb and 2 kb regions could be amplified), except for

amplification of the full length 3.3 kb construct, which required at least 10 ng plasmid DNA to obtain a faint 3.3 kb band (Figure-A.24, lane 8).

Due to the difficulties in amplifying DNA fragments larger than 2kb, I designed primers, PC1 and PC2, that would amplify from just outside of the HI0938 gene (55 bp upstream of the start codon, and 27 bp downstream of the stop codon) resulting in a 593 bp product for the wild type HI0938, or 1219 bp for the KAN replacement of HI0938. The PC1 and PC2 primers should definitively identify ΔHI0938 replacement mutants versus wild type strains, without the difficulties of amplifying the longer DNA fragments.

To test the efficacy of PC1 and PC2 primers, I used PCR with serially diluted KW20 and pLKR DNA templates (Figure-A.25). In parallel, PCR was done using K1 and K2 primers for the KAN fragment. Positive and negative controls worked as expected (Figure-A.25, lanes 1 to 7). Colony PCR with the RR1172 strain (Figure-A.25, lanes 8 and 13) did not work for either the PC1/PC2 or KAN K1/K2 amplicons, which highlights the unreliability with using colony PCR (i.e. amplification with KAN primers K1/K2 usually work). However, using chromosomal DNA as template, the putative ΔHI0938 strain RR1172 was clearly found to have HI0938 replaced with KAN (Figure-A.25, lane 9). The amplification yield with RR1172 DNA template was much less than that obtained using a comparable amount of KW20 DNA template. This indicates a possible mutational change in RR1172 where PC1 or PC2 anneals to during PCR.

To test which primer annealing site was affected (i.e. PC1 or PC2), different primer sets using PC1 and PC2 with either K1 or K2 were used (Figure-A.26). The positive controls

(593 bp band using KW20 DNA or colonies; and 1219 bp band with pLKR DNA) and negative controls (no amplification) worked as expected (Figure-A.26, lanes 1 to 6, and 15 to 19). The PC1 primer was identified as having difficulties annealing to RR1172 chromosomal DNA during PCR, especially when colonies were used as the DNA template source.

An attempt was made to confirm the genotype of RR1172 using purified chromosomal DNA, rather than colonies, as the template source (Figure-A.27). As in Figure-A.24, when colonies were used as the template source, the 1 kb bands (KAN, LEFT, and RIGHT) could be amplified from chromosomal RR1172 DNA (Figure-A.27, lanes 16, 17 and 18). A faint 1219 bp band with PC1 and PC2 primers (Figure-A.27, lane 15), and a 1085 bp band with PC1 and K2 primers (Figure-A.27, lane 20) were observed. Although the ~2 kb regions using P1/K2 and K1/P4 did not amplify from RR1172 DNA (Figure-A.27, lanes 19 and 21), the results with PC1/PC2 primers indicate that RR1172 is not wild type for HI0938, and therefore has the HI0938 gene replaced with the KAN insert.

An independent transformation experiment was done to generate replicate ΔHI0938 strains, since RR1172 was the only clone obtained from the initial screening process, and apparently contains a genetic alteration (i.e. in the LEFT region) other than simply the replacement of HI0938 with KAN. As with the first attempt, similar colony sizes were observed in the second trial (i.e. on kanamycin selective plates, numerous microcolonies and several large colonies were present). Thus, microcolony pools were screened to

isolate HI0938 deletion mutants. Ten colonies were screened from one pool, and four of these were found to contain the KAN insert (Figure-A.28).

Chromosomal DNA from these clones, were tested further, using primer sets PC1/PC2, P1/K2, and K1/P4, to confirm the replacement of HI0938 with KAN (Figure-A.29). Only one of the putative ΔHI0938 strains (RR938) had the desired amplification for all three primer sets used (Figure-A.29, lanes 10 to 12). This demonstrates that the RR938 strain has the HI0938 gene replaced by KAN. In addition, PCR revealed that the other three clones were not wild type for the HI0938 gene. All four ΔHI0938 strains (i.e. RR1172, RR938, P3-1, P4-1, and P4-2) were checked for competence defects (described in Chapter 4).

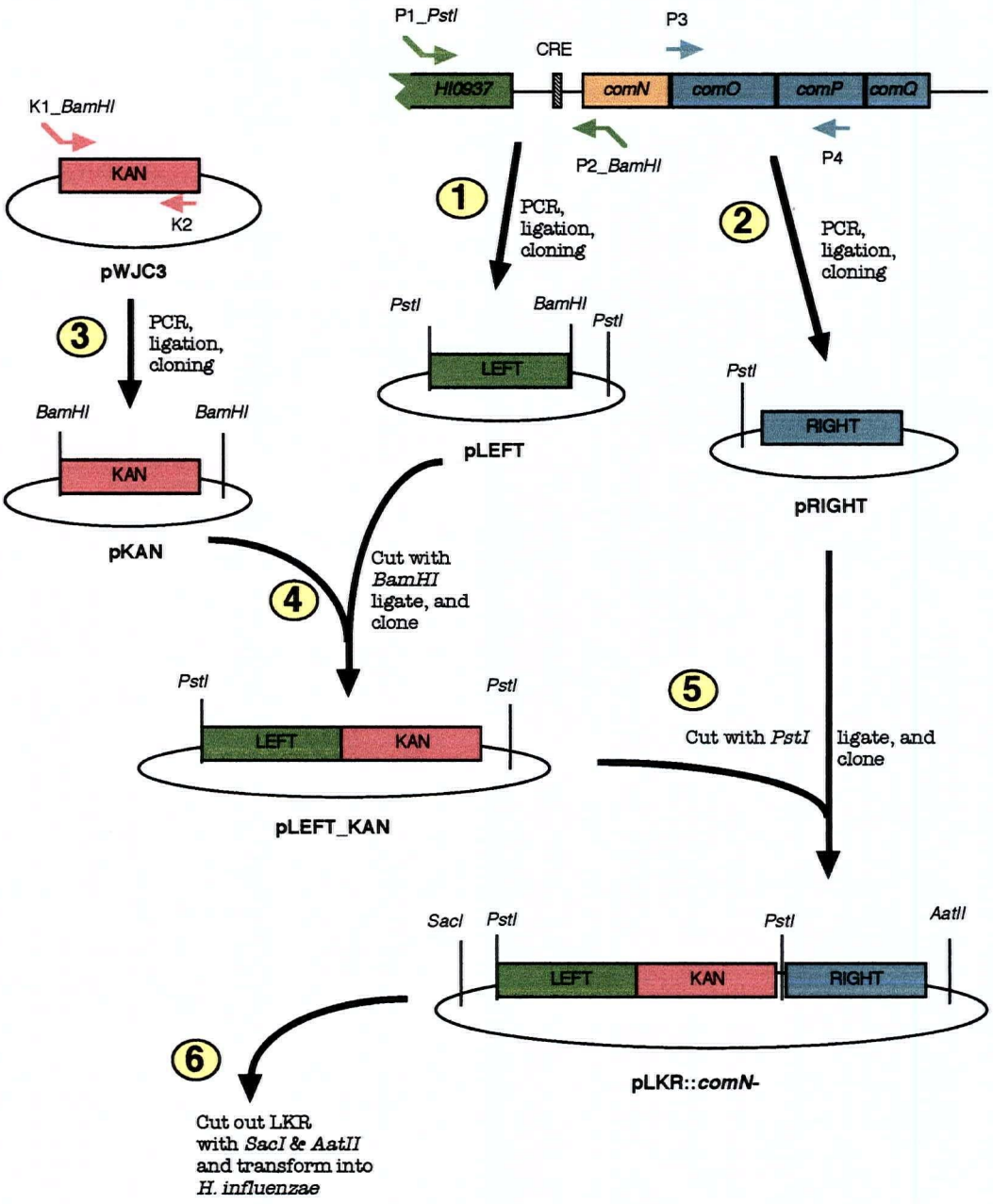


Figure-A.1: Overall scheme for *comN* (HI0938) knockout construction

This figure is reproduced from Figure-1 of Chapter 3 for convenience. CRE, Competence Regulatory Element; LKR, 3.3 kb LEFT_KAN_RIGHT construct.

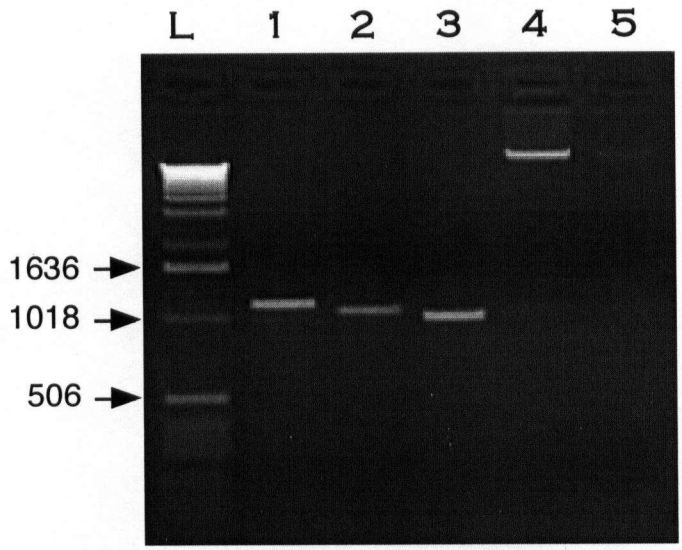


Figure-A.2: Gel purified LEFT, RIGHT and KAN inserts. [EXPT#064-4]

The three DNA fragments (LEFT, RIGHT and KAN) were each amplified by PCR, then purified by electrophoresis and gel extraction. DNA concentrations calculated from the absorbance readings with a UV spectrophotometer were 170 ng/ul for LEFT and KAN, and 150 ng/ul for RIGHT. Lane **L** is the 1 kb ladder. Lanes **2**, **3** and **4** are gel purified PCR products of LEFT, RIGHT, and KAN, respectively, and were run in this gel to determine DNA quality and approximate quantities after use of the gel extraction kit. Lanes **5** and **6** are 100 ng and 10 ng, respectively, of MAP7 DNA as a comparison of band intensity for DNA amounts loaded in the gel. The gel purified PCR products were ligated directly into pGEM-T Easy or pTOPO (Figure-A.1, steps 1, 2, and 3) and then cloned into *E. coli*.

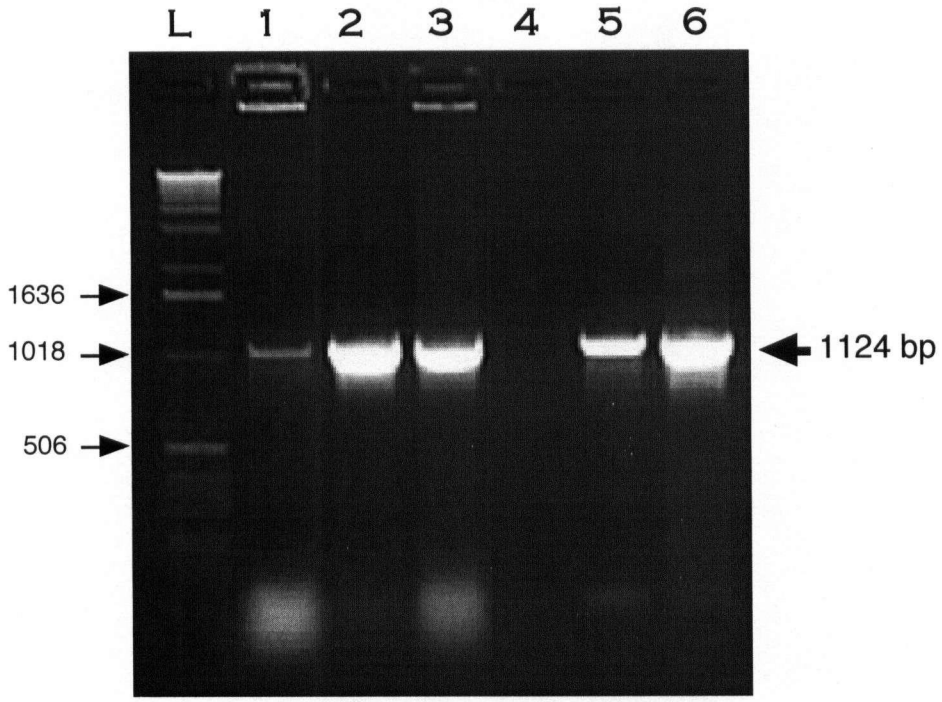


Figure-A.3: Colony PCR shows the presence of the LEFT insert in a pLEFT *E. coli* transformant. [EXPT#069-3]

After transformation of ligation reactions of the LEFT insertion into pGEM-T Easy vector, colony PCR was done to screen clones containing the pLEFT construct. Ampicillin resistant clones were chosen based on blue/white screening. Lane **L** is the 1 kb ladder. Lane **1** is the LEFT PCR product from using a single colony as the DNA template source. Lanes **2** and **3** are PCR products done on samples from two independent overnight cultures inoculated from the same colony used for lane **1**. The bands in lanes **2** and **3** are higher in product yield than the band in lane **1**, because the overnight cultures used for the DNA source in PCR were at a much higher cell density (and therefore, contained a higher amount of template DNA) than for the single colony PCR. Lane **4** is a negative control PCR using DH5α DNA as the template (i.e. no band expected). Lanes **5** and **6** are positive controls using the templates KW20 DNA and purified LEFT DNA, respectively. The bands present are of the expected 1124 bp size for the LEFT product. Minipreps of this pLEFT clone (strain #L-1B-1; RR1160) were then used for restriction digest analysis (Figures-A.5 and –A.6).

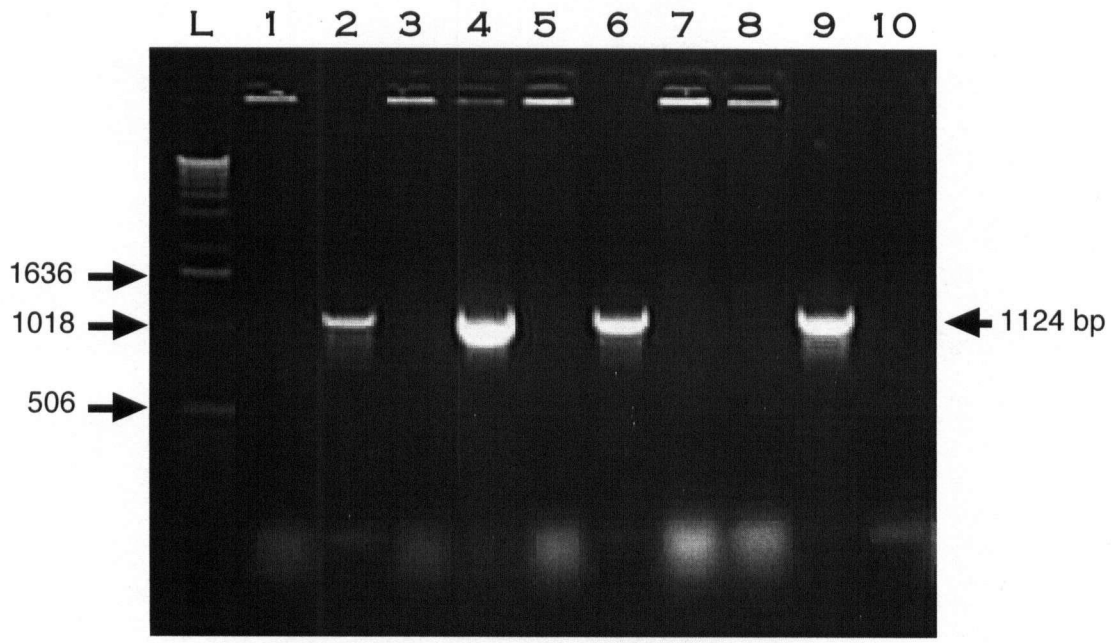
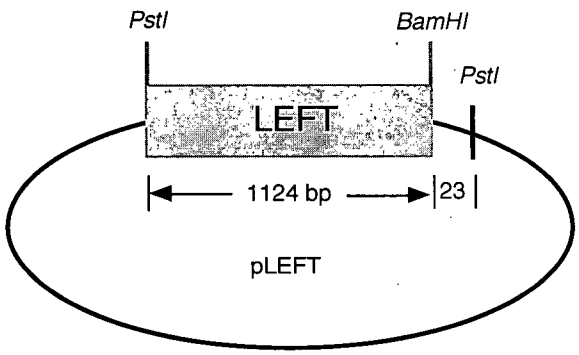


Figure-A.4: Screening for pLEFT clones using colony PCR. [EXPT#074-3]

In addition to the pLEFT clone from figure-A.3, eight more clones chosen with blue/white screening from an independent transformation experiment were tested for the presence of the LEFT insert using colony PCR. Of the eight, only three clones were found to have the expected 1124 bp LEFT insert (lanes **2**, **4** and **6**, for clones L-2, L-4, and L-6). Lane **9** is the positive control PCR using KW20 DNA as the template (1124 bp band expected), and lane **10** is the negative control using DH5α DNA (no band expected). Minipreps of the three clones with LEFT insert were then used for restriction digest analyses (Figures-A.5 and –A.6).

(A) Correct Orientation:

Expected sizes (*PstI*):
1147 bp (LEFT + Vector)
2993 bp (Vector)



(B) Reverse Orientation:

Expected sizes (*PstI*):
4117 bp (LEFT + Vector)
23 bp (Vector)

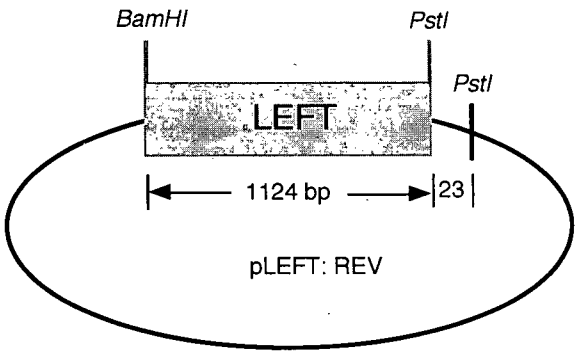


Figure-A.5: Restriction map of expected pLEFT correct and reverse orientations.

The LEFT fragment inserted into pGEM-T Easy can occur in either of two orientations. The desired orientation is that in (**A**). Differentiating between the two orientations can be easily done through *PstI* restriction digests, as a *PstI* site is present on the vector and another *PstI* site was engineered to occur at the end of the LEFT insert.

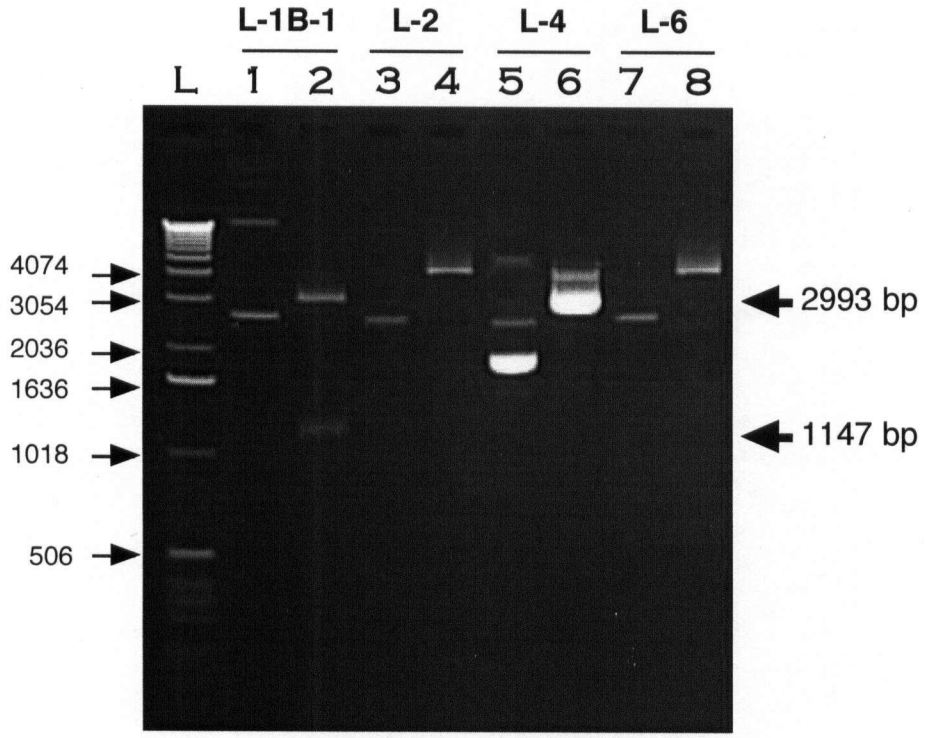


Figure-A.6: Restriction digests using PstI on pLEFT plasmids. [EXPT#075-3]

Restriction digests with PstI were used to confirm the presence, and to determine the orientations, of LEFT inserts in pLEFT plasmids. For the expected band sizes of digests with the correct and reverse orientations, see figure-A.5. Lane **L** is the 1 kb ladder. Lanes **1** and **2** are, respectively, the undigested and *PstI* digested pLEFT isolate #L-1B-1 (RR1160) from figure-A.3. Undigested and *PstI* digested plasmids for isolates #L-2 (lanes **3** and **4**), #L-4 (lanes **5** and **6**), and #L-6 (lanes **7** and **8**) are also shown. Plasmid from RR1160 clearly has the LEFT insert with the correct orientation (expected band sizes of 1147 bp and 2993 bp vector), while isolates L-2 and L-6 have the LEFT insert in the reverse orientation (expected size of 4117 bp). It is, however, unclear what the constitution of the pL-4 plasmid is. The pLEFT plasmid from RR1160 (#L-1B-1) was used to continue with the knockout procedure of HI0938 (Figure-A.1, step 4).

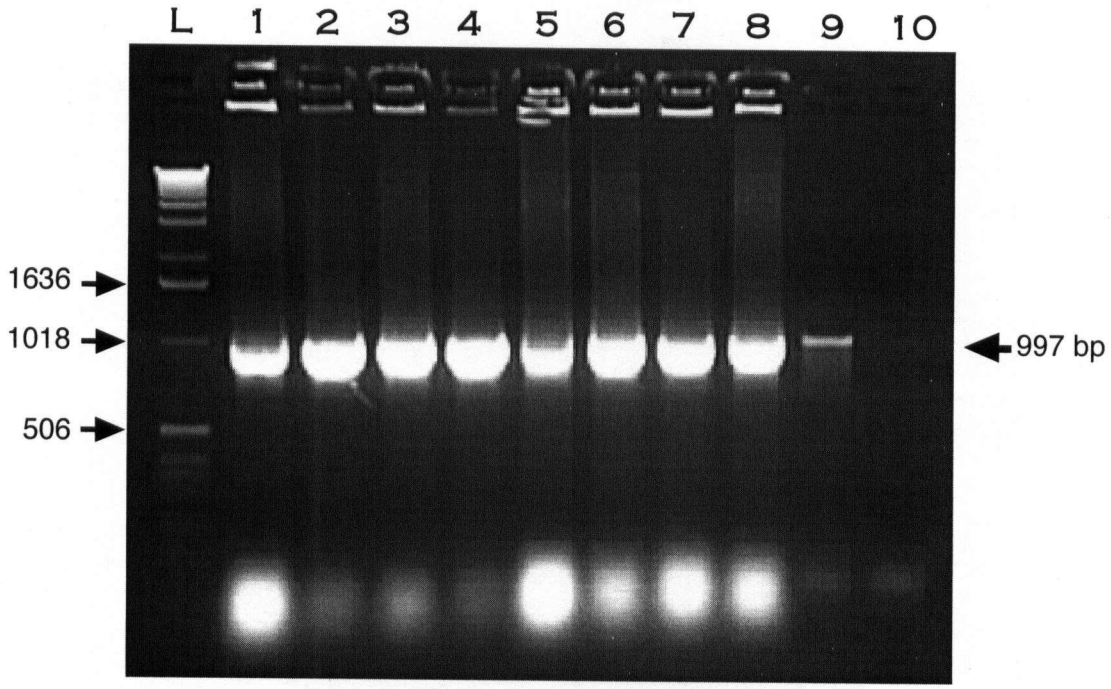
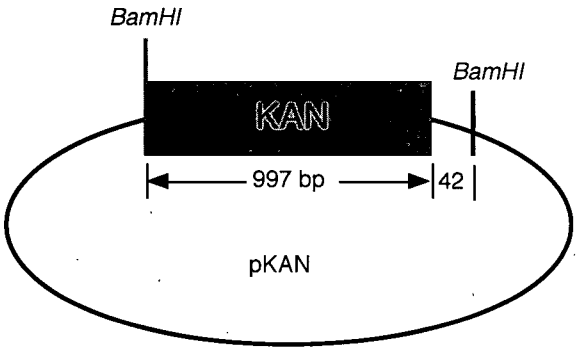


Figure-A.7: Colony PCR to screen for pKAN clones. [EXPT#077-2]

After transformations with ligations of the KAN insert into pTOPO vector, colony PCR was used to screen clones for pKAN. A total of sixteen colonies doubly resistant for ampicillin and kanamycin were chosen using blue/white screening. *BamHI* restriction digests of plasmids from the initial eight clones tested indicated that these plasmids had the KAN cassette in the reverse orientation (not shown). Thus, eight more clones were tested for the desired pKAN construct. Lane **L** is the 1 kb ladder. Lanes **1** to **8** are colony PCRs for the second set of eight colonies that were screened. Lane **9** is the positive control using pWJC3 DNA as the template (997 bp band expected). Lane **10** is the negative control with DH5α chromosomal DNA as the template (no band expected). The KAN insert was detected for the clones tested and for the positive control. Minipreps were then done on overnight cultures from these eight clones in preparation for restriction digest analysis (Figures-A.8 and –A.9).

(A) Correct Orientation:

Expected sizes (*BamHI*):
1039 bp (KAN + Vector)
3858 bp (Vector)



(B) Reverse Orientation:

Expected sizes (*BamHI*):
4897 bp (KAN + Vector)
42 bp (Vector)

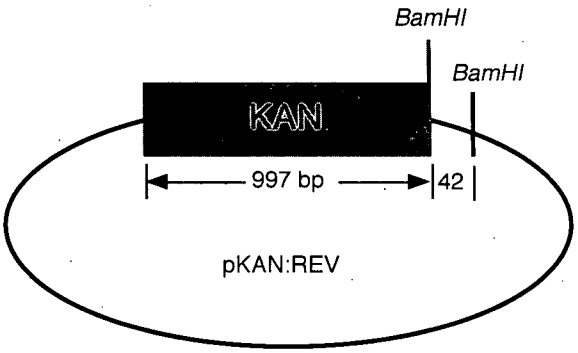


Figure-A.8: Restriction Map of Expected pKAN Correct and Reverse Orientations.

The RIGHT fragment inserted into the pCR2.1 TOPO TA vector can occur in one of two orientations. The desired orientation is that in (**A**). Use of the engineered *BamHI* site on the insert, and the *BamHI* site already present in the vector, allow for distinguishing different pKAN plasmids with the two possible KAN insert orientations.

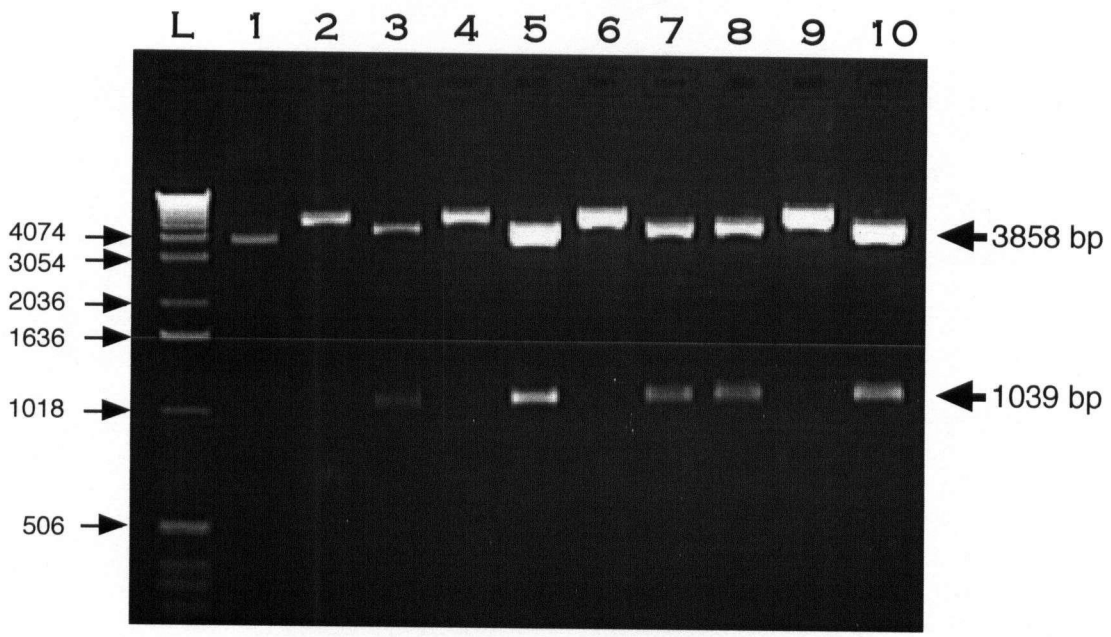


Figure-A.9: *BamHI* Restriction digest analysis of pKAN plasmids. [EXPT#077-5]

Minipreps of each pKAN clone were prepared for *BamHI* restriction digest analysis. See figure-A.8 for the expected band sizes of the correct and reverse KAN insert orientations. Lane **L** is the 1 kb ladder. Lane **1** is a representative undigested pKAN plasmid. Lane **2** is a positive control for the *BamHI* digestion on a plasmid with KAN in the reverse orientation (pKAN:REV, expected size of 4897 bp). Lanes **3** (clone #K-9), **5** (K-11), **7** (K-13), **8** (K-14) and **10** (K-16) show digests of pKAN with the correct orientation (expected band sizes of 1039 bp and 3858 bp). Lanes **4**, **6** and **9** are pKAN:REV plasmids (expected size of 4897 bp; compare with lane **2**). One of the pKAN plasmids (RR1162, lane **5**) was *BamHI* digested and purified for the insertion of KAN into pLEFT (Figure-A.1, step 4, and Figure-A.13).

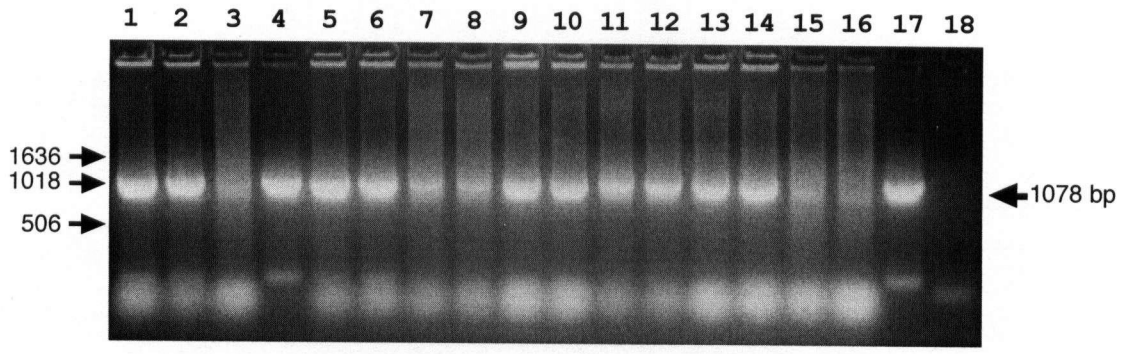
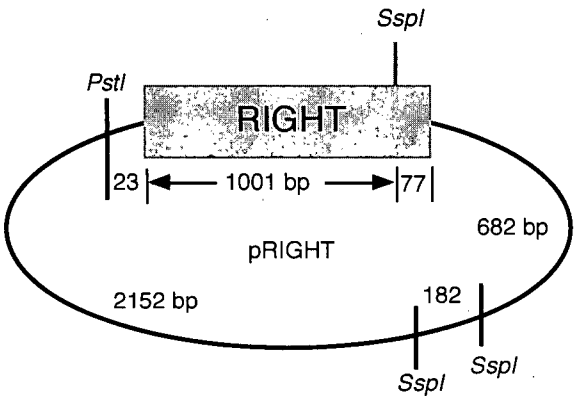


Figure-A.10: Colony PCR to screen for pRIGHT clones. [EXPT#092-7]

After transformation with ligation reactions of the RIGHT insert into pGEM-T Easy vector, colony PCR was done to screen clones for the pRIGHT construct. Ampicillin resistant clones were chosen based on blue/white screening. Lanes **1** to **16** show colony PCR for different clones to detect the presence or absence of the RIGHT insert. Colonies for lanes **3**, **4**, **7**, **8**, **11**, **12**, **15**, and **16** correspond to clones #R-3, -4, -7, -11, -12, -15, and -16, respectively. Lane **17** is a positive control using the gel purified RIGHT insert as a DNA template for PCR. Lane **18** is a negative control using DH5α DNA as the template. Minipreps were done on all overnight cultures of clones (except R-3, R-7, R-15 and R-16), and were analysed for insert orientation using *SspI* restriction digests (Figures-A.11 and –A.12).

(A) Correct Orientation:

Expected sizes (*SspI*):
3153 bp (RIGHT + Vector)
759 bp (RIGHT + Vector)
182 bp (Vector)



(B) Reverse Orientation:

Expected sizes (*SspI*):
2222 bp (RIGHT + Vector)
1683 bp (RIGHT + Vector)
182 bp (Vector)

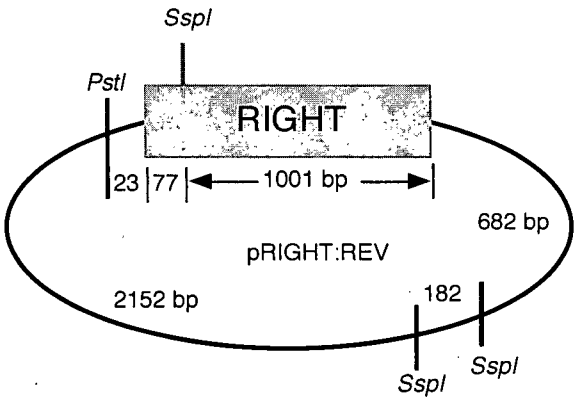


Figure-A.11: Restriction map of expected pRIGHT correct and reverse orientations.

The RIGHT fragment inserted into pGEM-T Easy can occur in one of two orientations. The desired orientation is that in (**A**). *SspI* was identified as being the best diagnostic restriction enzyme for determining the orientation of RIGHT.

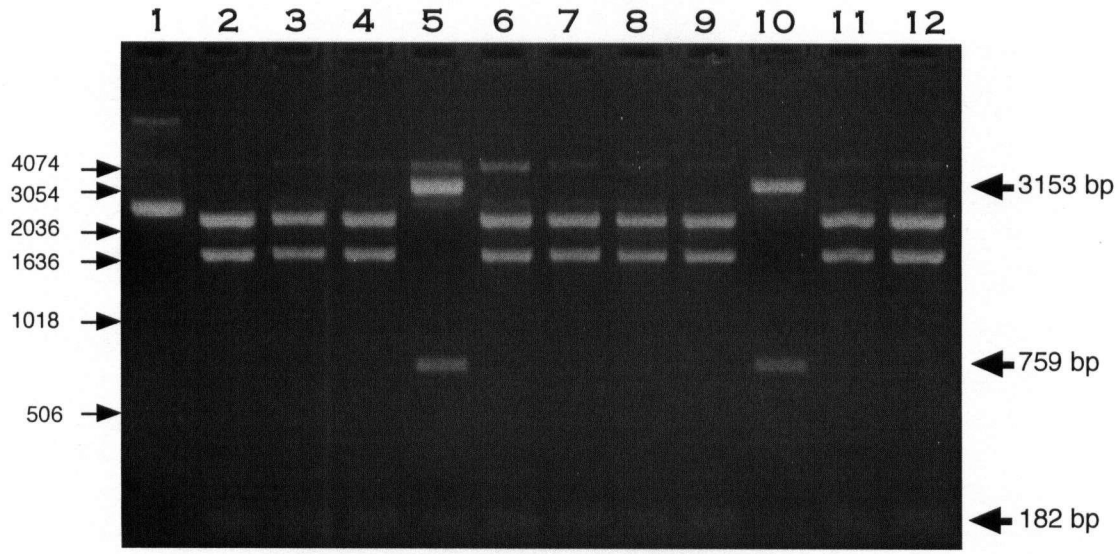


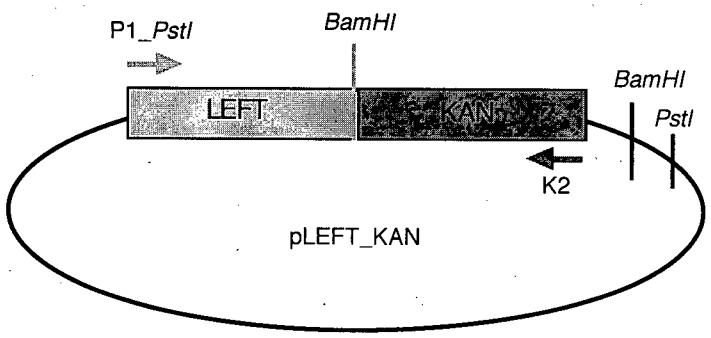
Figure-A.12: Restriction digests using SspI on pRIGHT plasmids. [EXPT#092-9]

Restriction digests on pRIGHT plasmids with the *SspI* enzyme was done to determine the RIGHT insert orientation in pGEM-T Easy vector. The expected band sizes for correct and reverse orientations are given in figure-A.11. Lane **1** is a representative undigested pRIGHT plasmid. Lane **2** is positive control *SspI* digested pRIGHT:REV plasmid. Lanes **3** to **12** are *SspI* digests of pRIGHT plasmids from figure-A.7. The band near 4074 bp is the single cut (i.e. partial digest) of pRIGHT (4094 bp). The 182 bp band was present in all digested plasmids as expected. Only two pRIGHT plasmids were found to contain RIGHT in the desired orientation (lanes **5** and **10**, clones R-4 and R-11, respectively), based on the expected band sizes of 3153 bp, 759 bp, and 182 bp that were present. One of the clones (RR1161; clone #R-4, lane **5**) was used to construct the pLKR plasmid (Figure-A.1, step 5).

(A) Correct Orientation:

Expected size (P1_*PstI* & K2):

2121 bp LEFT_KAN



(B) Reverse Orientation:

Expected size (P1_*PstI* & K2):

no bands expected

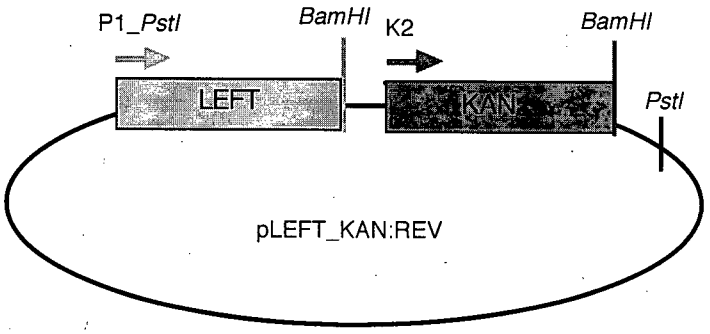


Figure-A.13: Expected PCR products for pLEFT_KAN.

When primers P1_*PstI* and K2 are used on pLEFT_KAN, with KAN in the correct orientation, a 2121 bp PCR product is expected. Conversely, no PCR products are expected when KAN is in the reverse orientation.

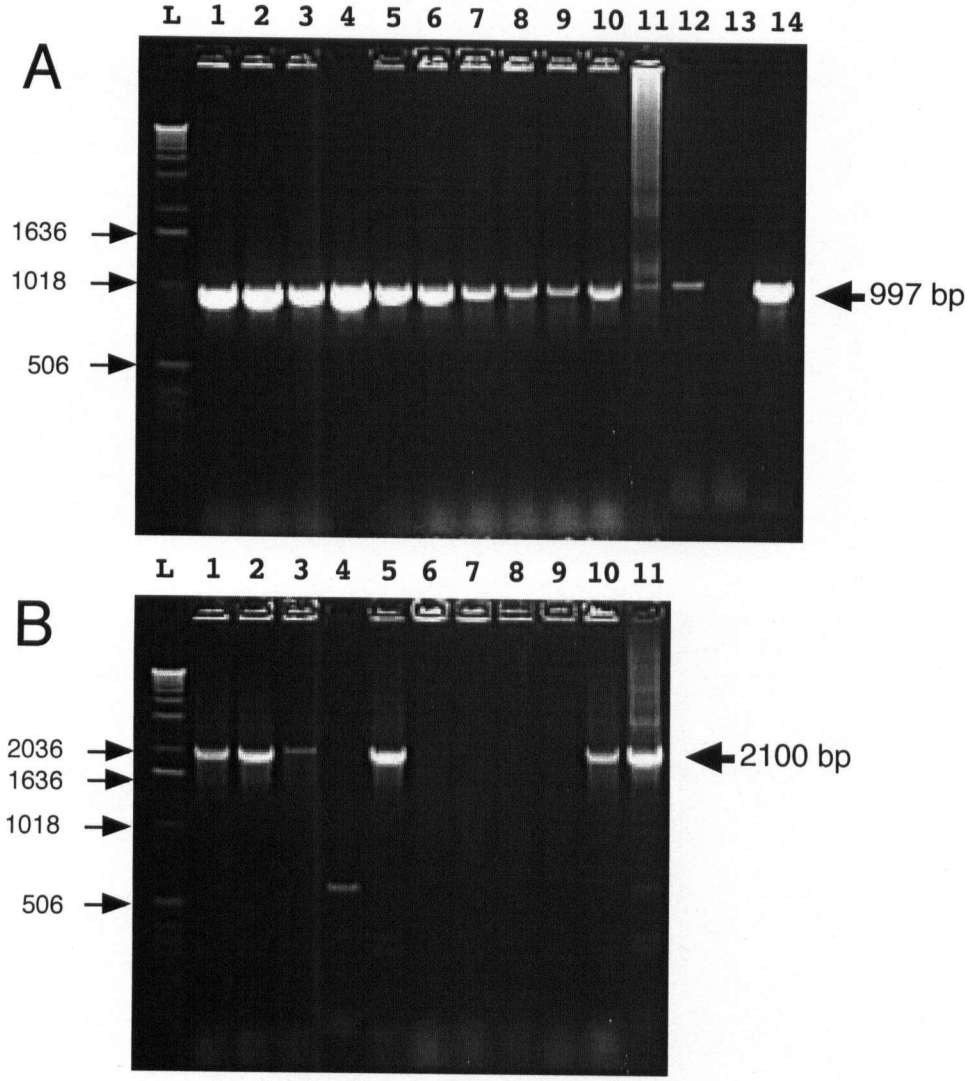
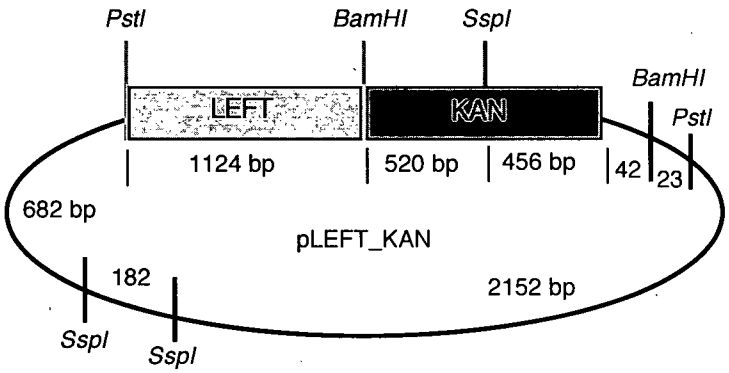


Figure-A.14: Colony PCR to screen for pLEFT_KAN clones. [EXPT079-4]

In step 4 (Figure-A.1), the KAN fragment was ligated into pLEFT vector at the *BamHI* site, and ligations were transformed into competent *E. coli*. Eleven ampicillin and kanamycin resistant colonies were then screened for the presence of the kanamycin cassette, and for the LEFT_KAN fragment, using colony PCR. (**A**) The KAN fragment was detected in all eleven colonies tested (using K1 and K2 primers). Colony PCR for clone KL-11 (lane **11**) resulted in a high molecular weight smear, the reasons for which are uncertain. (**B**) Colony PCR for the KAN_LEFT fragment (using primers P1 and K2) was done in parallel with **A** to determine the KAN insert-orientation. Samples from overnight cultures of pKAN_LEFT clones containing the desired KAN orientation (KL-1, -2, -5, and -10, from lanes **1**, **2**, **5** and **10**, respectively) and with the putative reverse orientation (KL-6 and KL-7, lanes **7** and **8**, respectively) were stored at –80°C and used for minipreps. Plasmid pKL-1 (RR1163; lane **1**) was used for step 5 (Figure-A.1) in the HI0938 knockout procedure.

(A) Correct Orientation:

Expected sizes (*SspI*):
2673 bp (LK + Vector)
2326 bp (K + Vector)
182 bp (Vector)



(B) Reverse Orientation:

Expected sizes (*SspI*):
2695 bp (LK + Vector)
2304 bp (K + Vector)
182 bp (Vector)

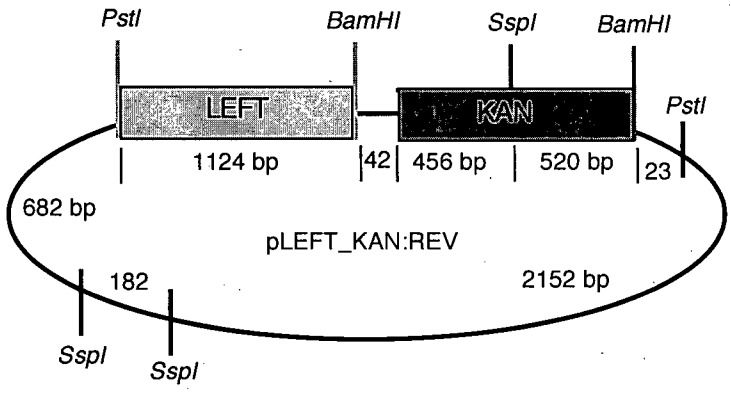


Figure-A.15: Restriction Map of Expected pLEFT_KAN Correct and Reverse Orientations.

The pLEFT_KAN (pLK) plasmid can occur with the KAN insert in one of two orientations. Distinguishing between the two pLK types could not be done with any of the available restriction enzyme sites, as DNA fragments from digests would be similar for either pLK or pLK:REV. Thus, the desired pLK construct was determined by using PCR with primers P1 and K2. Using the primers P1 and K1 should be able to detect plasmids with the reverse KAN orientation.

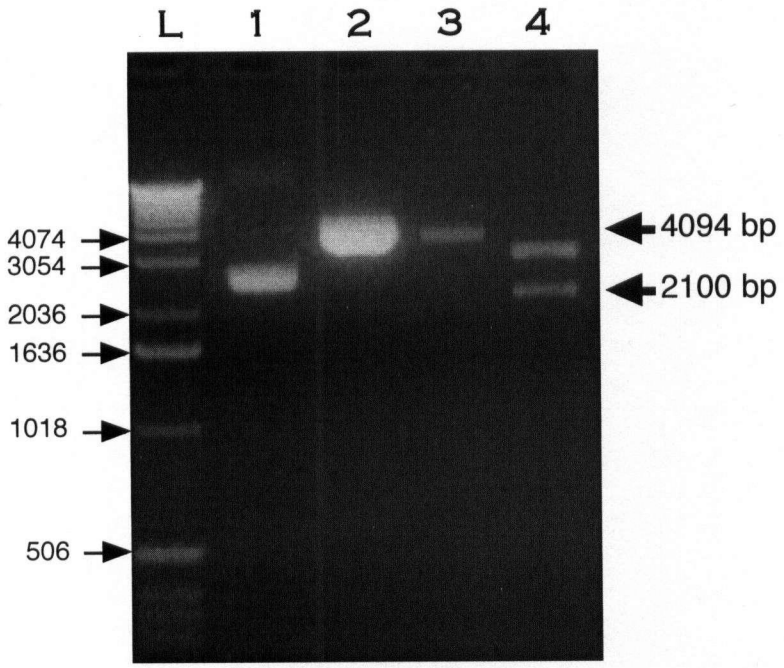


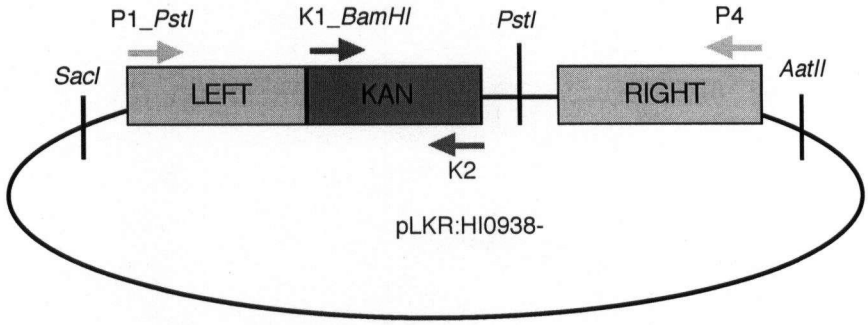
Figure-A.16: *PstI* digests of pRIGHT and pLEFT_KAN. [EXPT#079-6]

Plasmids pRIGHT and pLEFT_KAN were digested with *PstI*. The *PstI* digest of pLEFT_KAN had the expected bands of 2100 bp (LEFT_KAN insert) and 3016 bp (pGEM-T Easy vector). *PstI* digested pRIGHT had the expected 4094 bp band. Digests were purified before ligating the LEFT_KAN with pRIGHT for cloning the final pLKR construct into *E. coli*.

(A) Correct Orientation:

Expected sizes:

P1_*PstI* / K2	2121 bp LEFT_KAN
K1 / P4	2202 bp KAN_RIGHT



(B) Reverse Orientation:

Expected sizes:

P1_*PstI* / K2	2121 bp LEFT_KAN
K1 / P4	no bands

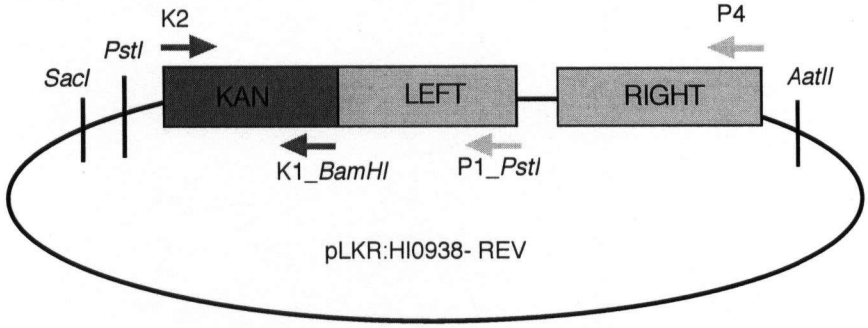


Figure-A.17: Expected PCR products for pLKR.

When P1 and K2 primers are used, a 2121 bp PCR product is expected for pLKR with LEFT_KAN in either orientation. When primers K1_*BamHI* and P4 are used, a 2202 bp band is expected when LEFT_KAN is in the desired orientation, while no bands are expected when LEFT_KAN is in the reverse orientation.

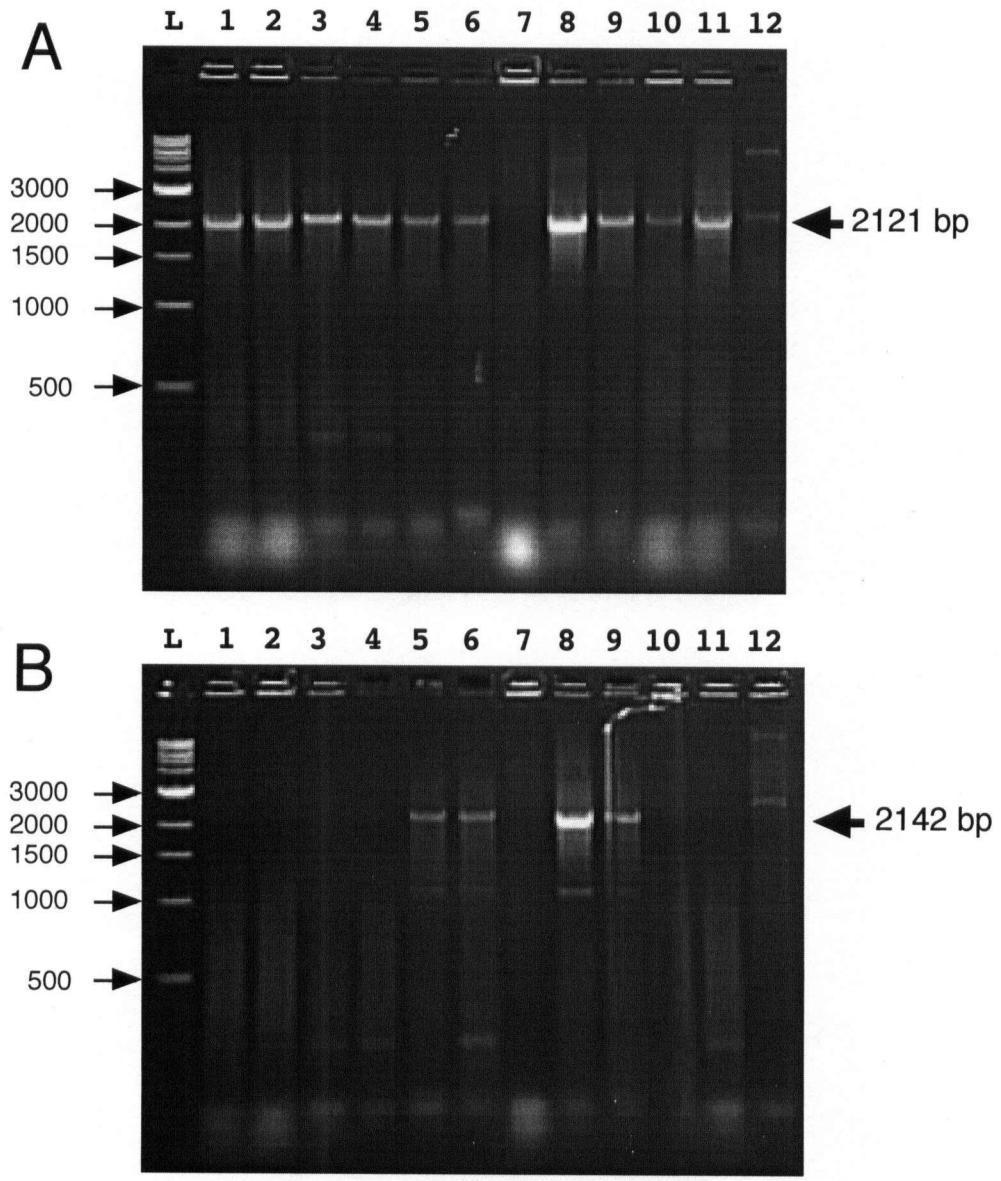
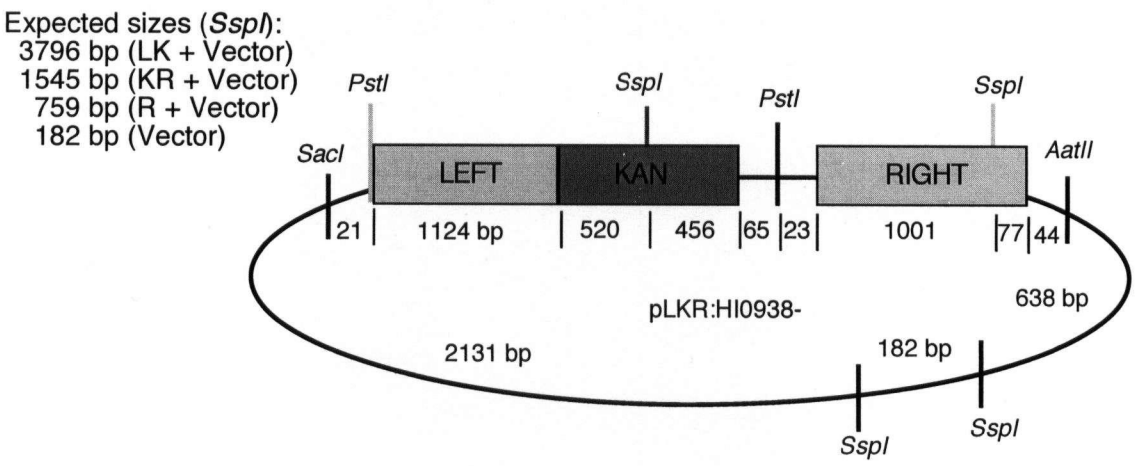


Figure-A.18: Colony PCR to screen for pLKR clones. [EXPT#095-14]

Ligations of the purified LEFT_KAN inserted into the *PstI* site of pRIGHT were transformed into *E. coli*. Colonies that were doubly resistant to ampicillin and kanamycin were then chosen for colony PCR to check for (**A**) the presence of the LEFT_KAN insert (using primers P1 and K2), and (**B**) for the correct orientation of LEFT_KAN (using primers K1 and P4). Except for one colony (lane **7**), all had the LEFT_KAN insert present, and four of the colonies tested had the LEFT_KAN insert in the correct orientation (pLKR-45, pLKR-46, pLKR-48, and pLKR-49, lanes **5**, **6**, **8**, and **9**). Samples from overnight cultures of these four pLKR clones (pLKR-45, -46, -48, and –49) were stored at -80°C, and minipreps were obtained for further analysis with restriction digests (Figures-A.19 and –A.20).

(A) Correct Orientation:



(B) Reverse Orientation:

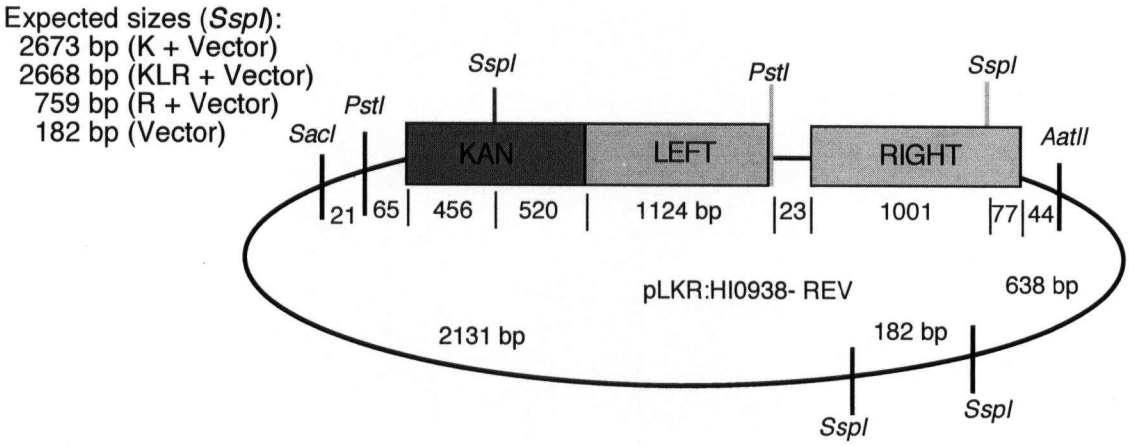


Figure-A.19: Restriction Map of Expected pLKR Correct and Reverse Orientations.

Once the LEFT_KAN insert is ligated into pRIGHT, the LEFT_KAN orientation can be determined by using *SspI* digests. The *SspI* positions on the plasmid are asymmetrical and therefore yield different DNA fragments for pLKR and pLKR:REV. The *PstI* sites can be used to confirm the presence of the LEFT_KAN insert (2165 bp for LEFT_KAN, and 4115 bp for the pRIGHT). In order to generate the ΔHI0938 strains, the *AatII* and *SacI* sites were used to cut out the 3.3 kb LKR construct to transform the linear DNA fragment into competent wild type KW20 cells.

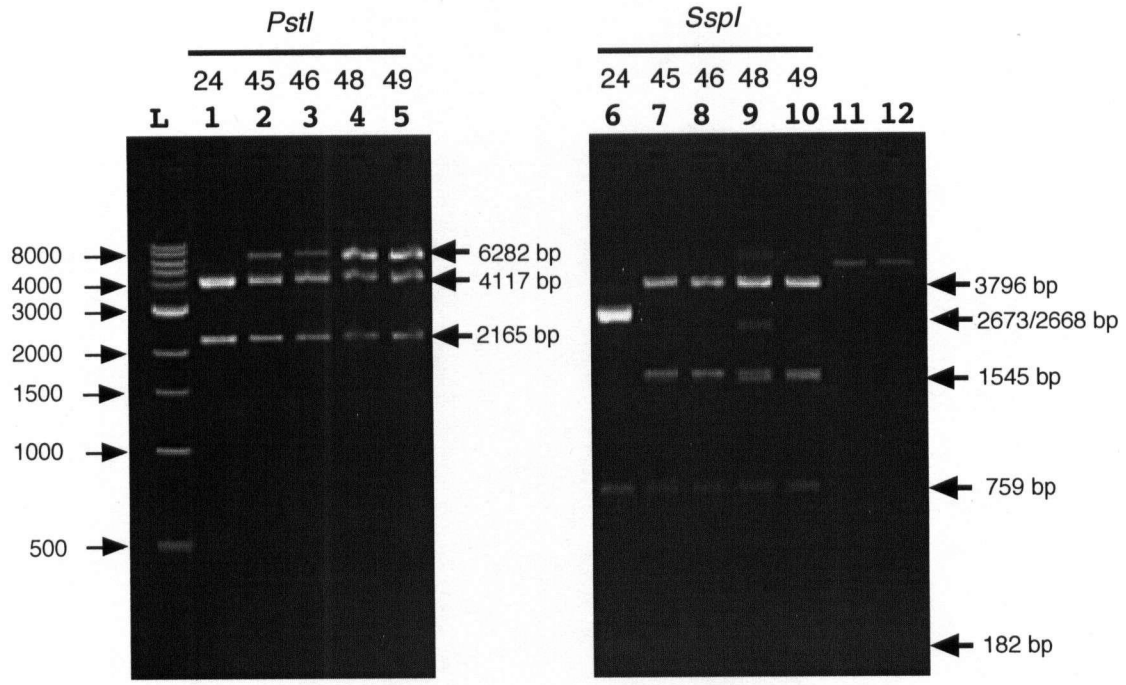


Figure-A.20: Restriction digests to confirm the correct orientations of LEFT_KAN in pLKR. [EXPT#095-21]

Minipreps of clones containing the desired pLKR plasmids (pLKR-45, -46, -48, and –49) were put through several different restriction digestions for confirmation of the correct LEFT_KAN insert orientation. *PstI* (lanes **1** to **5**) was used to confirm the presence of the LEFT_KAN (expected size of 2165 bp) insert in pRIGHT (4117 bp band). The 6282 bp band present in lanes **2** to **5** are partially digested (i.e. single cut) plasmid DNA. Lane **1** is a positive *PstI* digest control using a pLKR plasmid (clone #LKR-24) with the reverse LEFT_KAN orientation that was identified in a previous screen. *SspI* was found to be the best possible diagnostic restriction enzyme for the LEFT_KAN orientation in pLKR. See figure-A.19 for the expected band sizes of pLKR with LEFT_KAN in the correct and reverse orientations. Lane **6** is the *SspI* digest of the control pLKR-24 plasmid with LEFT_KAN in the reverse orientation (expected band sizes of 2673 and 2668 bp, 759 bp, and 182 bp). Lanes **7** to **10** are pLKR test plasmids (pLKR-45, -46, -48, and -49, respectively) digested with *SspI*. Three pLKR plasmids (pLKR-45, -46, and -49) are clearly the desired constructs. Plasmid pLKR-48 may also be the correct construct, but it was only partially digested and has a double band at approximately 1.5 kb, which suggests a potential sequence difference (e.g. mutation) compared to the other three plasmids. Lanes **11** and **12** are undigested plasmids pLKR-24:REV and pLKR-45. To be certain the pLKR plasmids were the desired constructs, an empirical analysis calculating band sizes from gel migration distances was done, and the results indicated that the constructs were as desired.

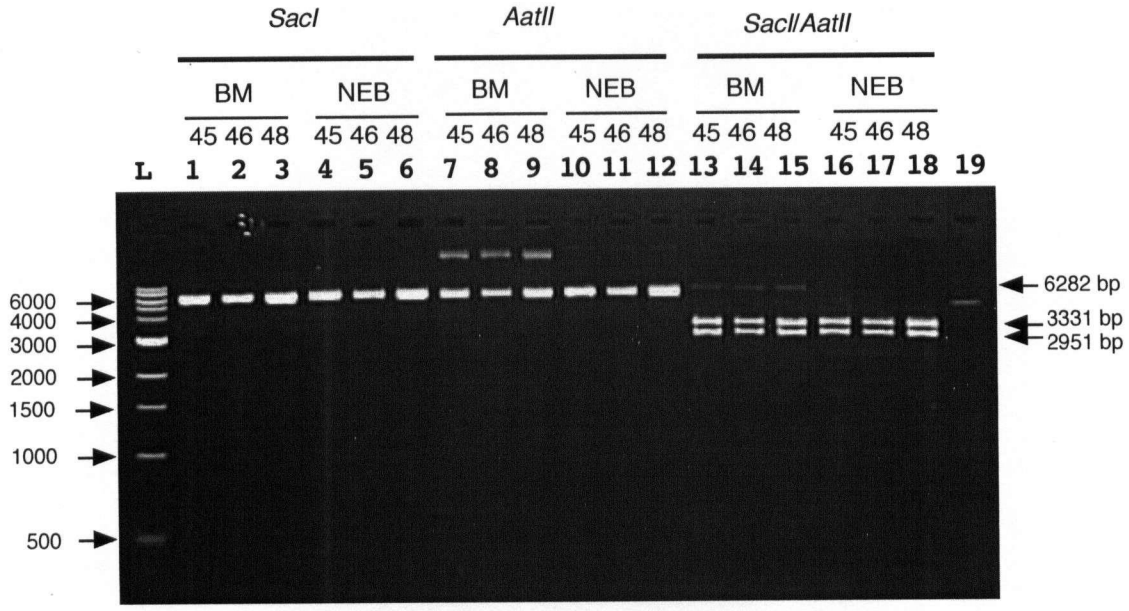


Figure-A.21: *AatII* and *SacI* double digests of pLKR plasmids. [EXPT#095-21]

Double digests using *AatII* and *SacI* were used to confirm the desired construction of pLKR plasmids (with the expected band sizes of 3331 bp for LKR construct and 2951 bp for the vector), and for transformation of the linear construct into *H. influenzae* to generate the HI0938 knockout (Figure-A.1, step 6). Three different pLKR plasmids (pLKR-45, -46, and -48) were each digested with *AatII* or *SacI* separately (lanes **1** to **12**). To determine optimal conditions for double digestions using *SacI* and *AatII* restriction enzymes (lanes **13** to **18**), two different reaction buffers were tested (New England Biolabs, NEB buffer#4 versus Boehringer Mannheim, BM Buffer L). The gel above was run for an additional 5 hours (approx.) to separate the *AatII*/*SacI* digests further, for a higher resolution of band sizes (not shown) to calculate the actual band sizes. Lane **19** is undigested plasmid pLKR-24 to show that the digests worked. Complete digestion was obtained with NEB buffer #4, while BM buffer L yielded a partial digestion. Unpurified *AatII*/*SacI* double digests with NEB buffer #4 were used for transformation of the linear LKR construct (from pLKR-45 or pLKR-46) into *H. influenzae* (Figure-A.1, step 6).

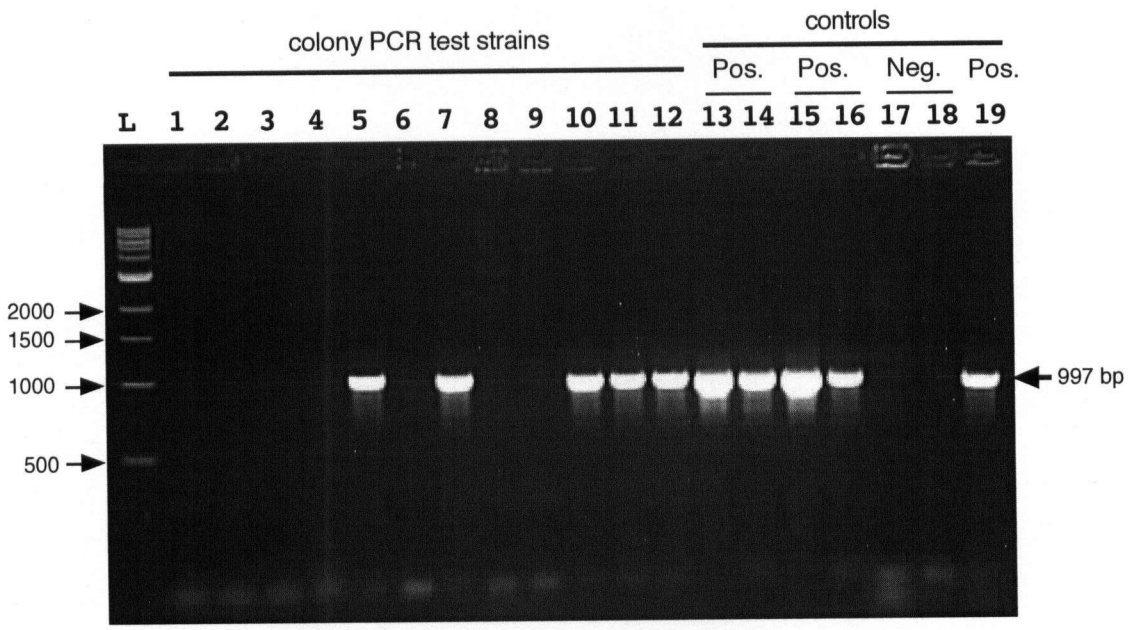


Figure-A.22: Colony PCR to screen pools of putative ΔHI0938 *H. influenzae* clones for the KAN insert. [EXPT#096-8]

Colony PCR was done to simultaneously screen for the KAN insert in seven pools of microcolonies (lanes 2, 3, 5, 7, 10, 11, and 12) and 5 normal sized single colonies (lanes 1, 4, 6, 8, and 9) that were putative ΔHI0938::*kan*^{R} *H. influenzae* transformants. Although none of the large colonies had the KAN insert, KAN was detected in five of the seven microcolony pools. Lanes 13 and 14 are positive PCR controls with approximately 1 ng and 0.1 ng of pLKR-46 undigested plasmid, respectively. Lanes 15 and 16 are positive controls using 1 ng and 0.1 ng of *AatII*/*SacI* digested pLKR-46 as plasmid DNA templates. Lanes 17 and 18 are negative controls using MAP7 DNA and a KW20 colony, respectively. Lane 19 is a positive control using a Δ*purR* colony. Colonies and pools had been streaked on plain and *kan* plates before colony PCR was carried out. To isolate pure cultures of putative ΔHI0938::*kan*^{R} strains, five colonies from streaked kanamycin plates for each of two pools (lane 5 and lane 10) were tested further with colony PCR for the presence of the KAN insert (Figure-A.23). These single colonies were also streaked on plain and *kan* plates, and were used to inoculate overnight cultures (both plain sBHI and kanamycin sBHI) in preparation for –80°C storage.

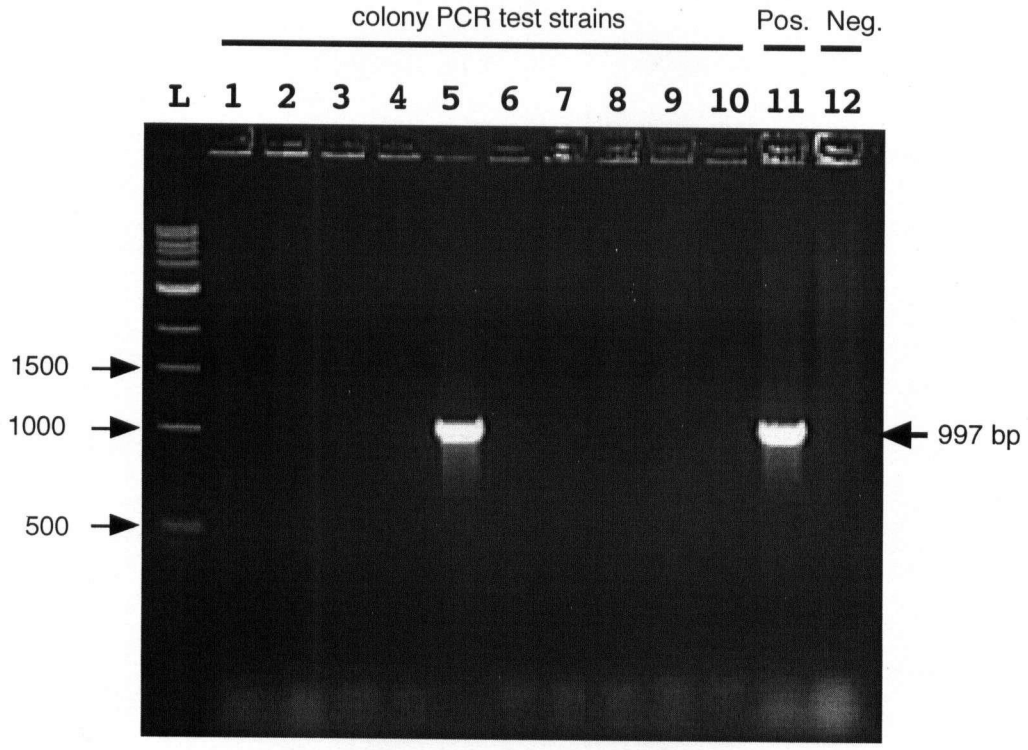


Figure-A.23: Colony PCR to screen individual putative ΔHI0938 *H. influenzae* clones for the KAN insert. [EXPT#096-12]

To determine whether the putative ΔHI0938 clones contain the KAN insert, colony PCR was done using primers K1 and K2. Lanes 1 to 5 are colony PCRs for five different colonies from microcolony pool number five, and lanes 6 to 12 are from pool number 10. Lane 11 is the positive control using a Δ*purR* colony, and lane 12 is a negative control using a KW20 colony. Although very faint bands of the expected KAN insert size (997 bp) could be seen in some of the lanes (e.g. lanes 6, and 7), the only definitive signal for the presence of KAN was in lane 5 (ΔHI0938 clone #5-5). Samples of this clone from overnight cultures were stored at –80°C. Colony PCR using different primer sets were the done to further characterize the genotype of ΔHI0938 #5-5, strain RR1172 (Figure-A.24).

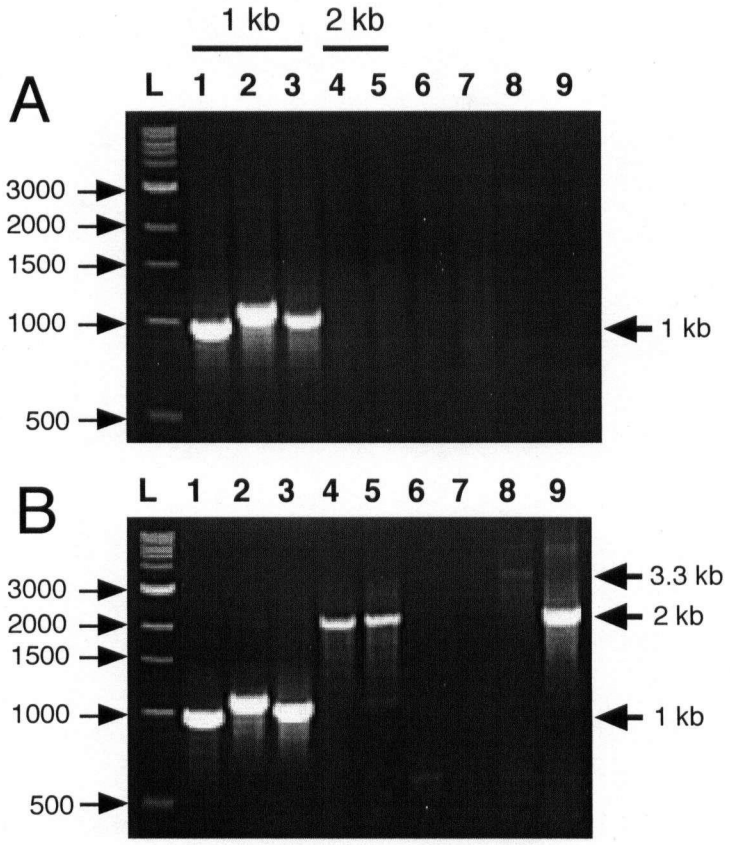


Figure-A.24: Colony PCR to determine genotype of the putative ΔHI0938 strain RR1172. [EXPT#096-14]

To characterize the genotype of the putative RR1172, colony PCR was done using a variety of primer sets, and the banding patterns were compared to PCR with pLKR DNA and KW20 chromosomal DNA templates. Lanes **1**, **2**, **3** in **A** and **B** were to amplify the KAN, LEFT and RIGHT DNA regions (each approximately 1 kb), and worked as expected using primer sets K1/K2 (997 bp), P1/P2 (1124 bp), and P3/P4 (1078 bp). Colony PCR to amplify 2.1/2.2 kb regions (e.g. LEFT_KAN or KAN_RIGHT, lanes **4** and **5**, respectively, with primer sets P1/K2 and K1/P4) did not work for RR1172 colonies (**A**), but did work in PCRs for pLKR DNA (**B**, lanes **4** and **5**). PCR to amplify the 3.3 kb region in ΔHI0938 colonies or the pLKR plasmid (lane **6**, in both **A** and **B**), or the 2.7 kb region in wild type KW20 DNA (lanes **8** and **9** in **A**) did not work as expected. The 3.3 kb amplification was not observed for RR1172 colonies, nor the 2.7 kb amplicon for KW20 DNA (using P1/P4 primers in both cases). Lane **8** in **B** did have a faint band at the expected 3.3 kb position, but only with pLKR plasmid at a rather high DNA concentration. A negative control for RR1172 and pLKR used primers (P2 and P4) for LK in the reverse orientation (lane **7** in **A** and **B**). Lane **9**, which used pLKR-24 (i.e. LK in the reverse orientation) as the DNA template in PCR with primers for the reverse LK orientation, worked as expected and confirms that the pLKR-46 plasmid (lanes **1** to **6**, **B**) is the desired construct (i.e. LK is not in the reverse orientation in pLKR-46).

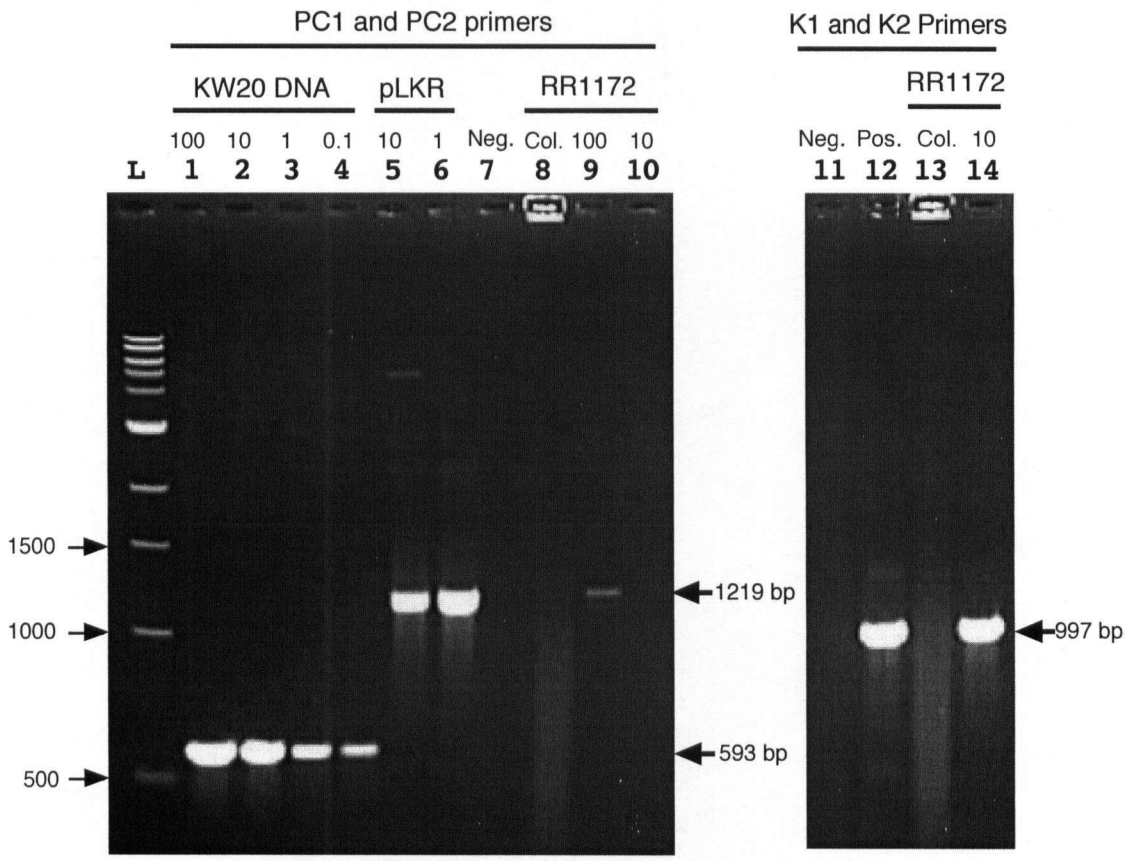


Figure-A.25: PCR to confirm the genotype of strain RR1172. [EXPT#101-3]

PCR was done on serially diluted KW20 chromosomal DNA (lanes **1** to **4**; numbers below bars are DNA amounts in nanograms) and on pLKR DNA (lanes **5** and **6**) to determine the efficacy of the PC1 and PC2 primers for diagnostic PCR. A negative control using *E. coli* DNA was included and worked as expected (i.e. no amplification in lane **7**). PCR was done on a RR1172 colony (lane **8**) and on purified chromosomal DNA (lanes **9** and **10**). A secondary set of PCR controls used K1 and K2 primers on the template DNA to amplify the KAN insert. Lane **11** is a negative control using KW20 DNA. Lane **12** is a positive control using pLKR-46 plasmid DNA as a template. Lanes **13** and **14** are PCRs using a RR1172 colony and purified DNA as templates, respectively. In this case, colony PCR did not work (lanes **8** and **13**), probably due to the age of the colonies, which highlights the unreliability of colony PCR. However, PCR did work on the purified DNA. A faint band in lane **9** shows that the strain RR1172 has the HI0938 gene replaced by the KAN insert. Confirmation of this result was done with different primer sets in diagnostic PCR (Figures-A.26 and –A.27).

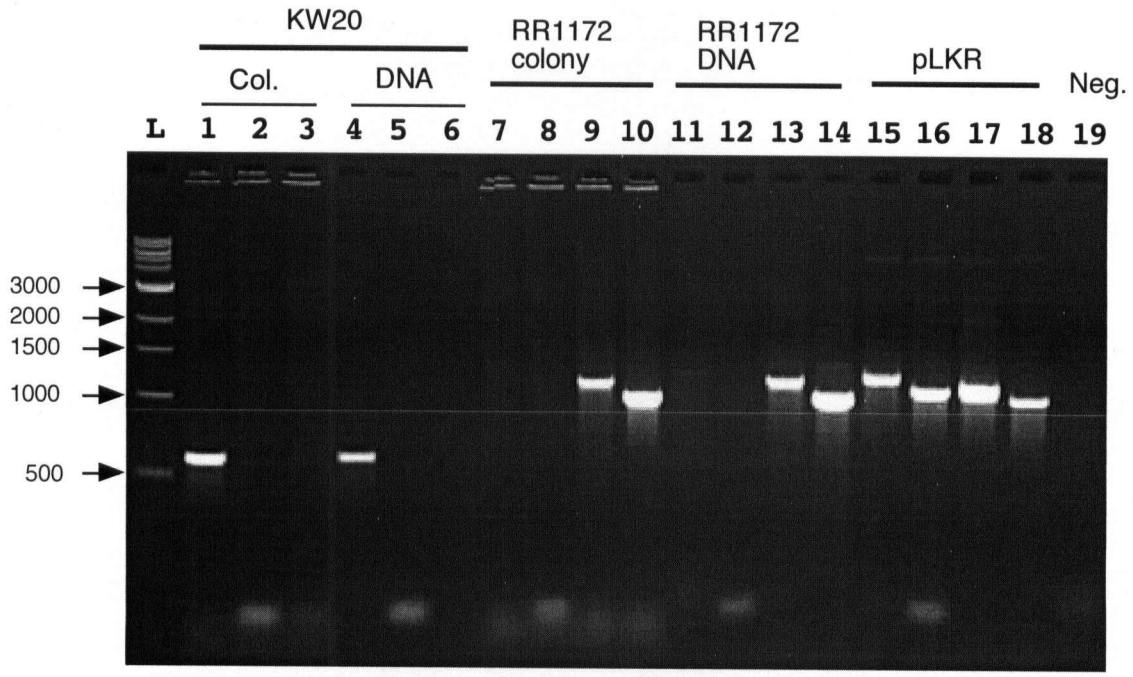


Figure-A.26: PCR using colonies and purified chromosomal DNA as templates to test PC1 and PC2 primers. [EXPT#105-4]

As PC1 and PC2 primers did not seem to work effectively for the RR1172 strain, as compared to the pLKR plasmid or KW20 DNA, a more rigorous approach in determining the problem was implemented. Colony PCR was compared to regular PCR (i.e. PCR done using purified DNA) using different primer sets on colonies and purified chromosomal DNA from KW20 and RR1172, and with PCR using the pLKR plasmid. Primers PC1 and PC2 were used on a KW20 colony and on DNA (lanes **1** and **6**, respectively), and the expected 593 bp amplicon was detected. The expected 1219 bp band was observed in PCR with pLKR (lane **15**), but not with the RR1172 colony (lane **7**). A faint band was detected when RR1172 purified chromosomal DNA was used as the template, however it was obviously not as strong a signal as obtained with pLKR DNA, or with wild type KW20 DNA. Negative controls included using primer sets PC1/K2 (lanes **2** and **5**) and K1/PC2 (lanes **3** and **6**), and K1/K2 (lane **19**) with a KW20 colony or chromosomal DNA as the template source. Positve controls for these primer sets were lanes **16** (PC1/K2), **17** (K1/PC2), and **18** (K1/K2) on pLKR DNA, with expected sizes of 1085 bp, 1131 bp and 997 bp, respectively. Primer set PC1/K2 did not work when either DNA or a colony from RR1172 was used as the template source, however strong signals were observed with primer sets K1/PC2 and K1/K2 were used. This indicates that the PC1 primer does not work effectively in the RR1172 strain, but works fine for wild type DNA or pLKR DNA. Further attempts to demonstrate whether the RR1172 strain had the HI0938 gene replaced by the KAN insert were carried out using purified DNA, rather than with colonies as the PCR template source (Figure-A.27).

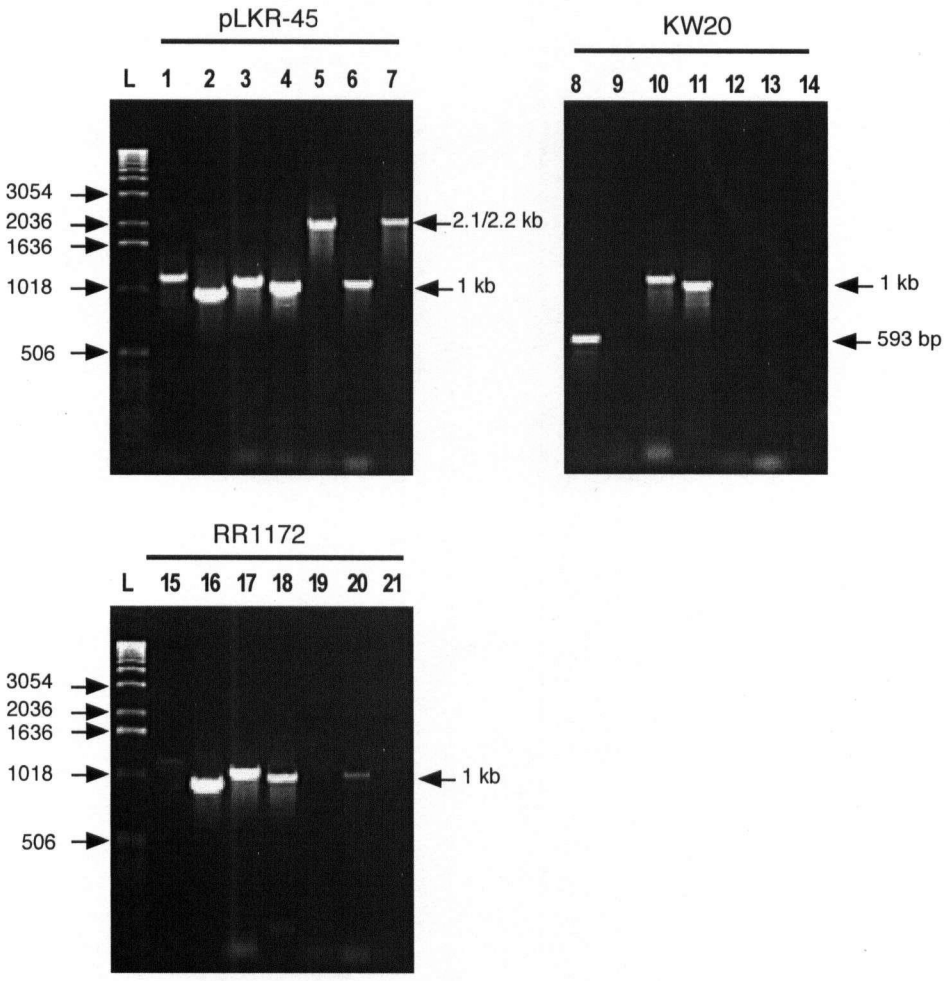


Figure-A.27: PCR with different primer combinations to confirm the RR1172 genotype. [EXPT#118-2]

PCR was carried out on purified DNAs to determine the genotype of RR1172. As in Figure-A.24, the ~1 kb regions (KAN, LEFT, and RIGHT) could be amplified from RR1172 chromosomal DNA and from the pLKR plasmid. The negative control for KAN (K1/K2 primers) on KW20 DNA worked as expected (i.e. no band in lane 9). The positive controls (lane **10**, P1/P2 primers for LEFT; lane **11**, P3/P4 primers for RIGHT) with KW20 DNA also worked as expected. However, attempts to amplify the 2.1 (P1/K2 primers for LEFT_KAN, lane 19) and 2.2 kb (K1/P4 primers for KAN_RIGHT, lane 21) regions from RR1172 DNA did not yield the desired product (i.e. no bands present). PC1/PC2 used in positive controls amplified the expected 593 bp fragment from KW20 (lane 8), and the 1219 bp from pLKR DNA (lane 1). A faint 1219 bp band (with PC1 and PC2) was present in lane **15** with RR1172 DNA template. Lane 20 shows a faint amplification using PC1 and K2 primers. Despite some difficulties in confirming the genotype of RR1172, probably due to some mutation in the LEFT DNA region that prevents PC1 from appropriately annealing during PCR, it seems likely that this strain is not the wild type form for the HI0938, and that it has the HI0938 gene replaced by the KAN insert.

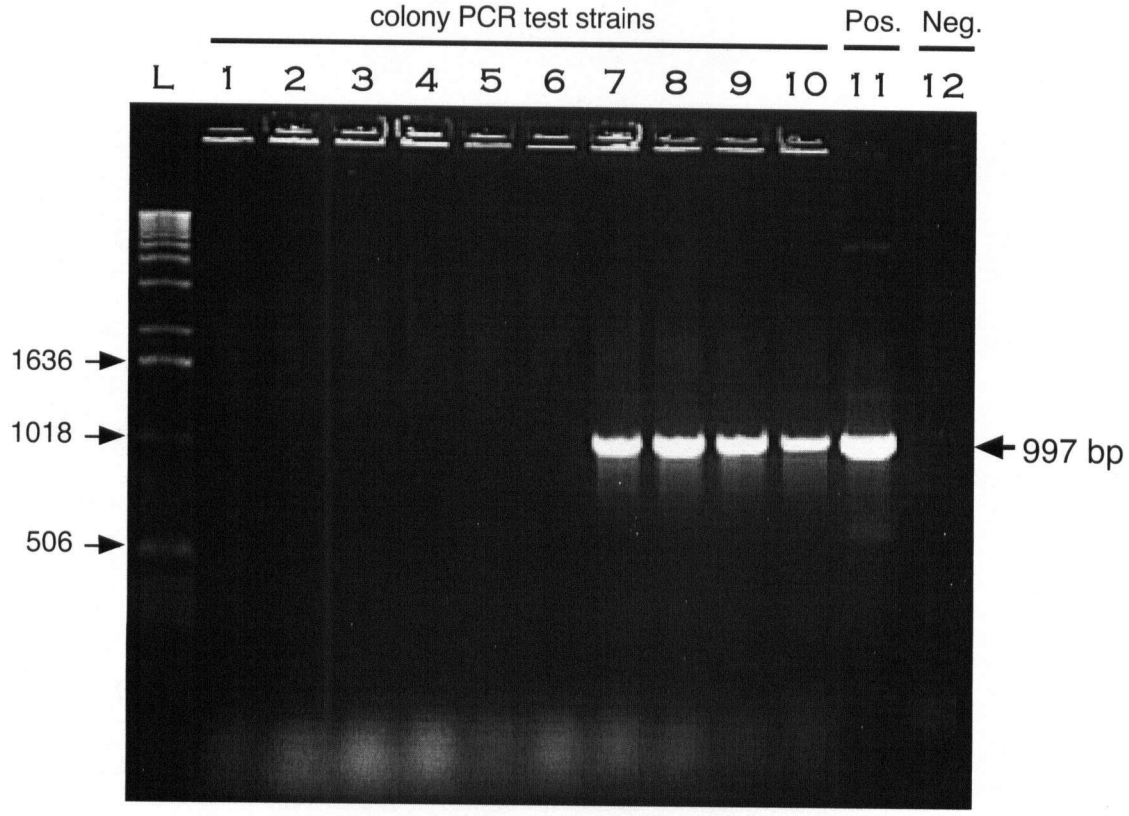


Figure-A.28: Screening for putative ΔHI0938 clones using colony PCR.
[EXPT#117-20]

From an independent replicate transformation experiment to generate Δ0938 strains, as that used to isolate clone RR1172, colony PCR was used to screen ten putative ΔHI0938::*kan*^{R} clones for the KAN insert (i.e. with K1 and K2 primers). The positive (lane **11**; 997 bp band) and negative (lane **12**; no band) controls worked as expected, using pLKR-45 and KW20 DNA as the templates, respectively. Four of the clones tested were found to contain KAN (lanes **7**, **8**, **9**, and **10**). Samples of these were stored at -80°C and used for chromosomal DNA extraction for further analysis (Figure-A.29).

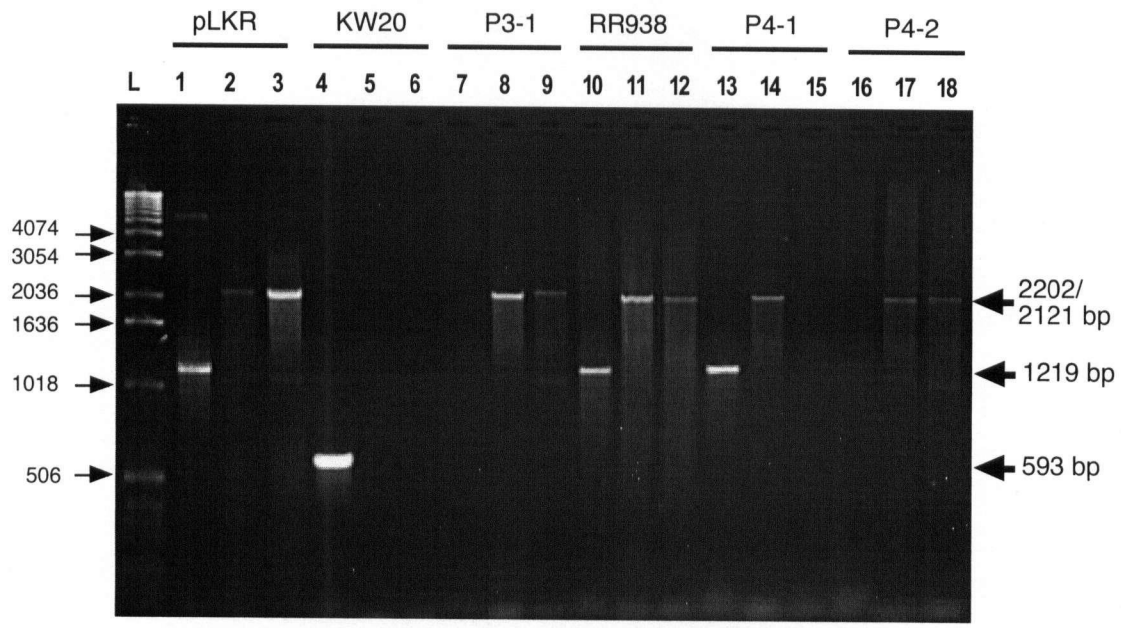


Figure-A.29: PCR on putative ΔHI0938 strains confirms that the HI0938 gene has been replaced by the KAN insert. [EXPT#119-8]

PCR was done on purified chromosomal DNA extracted from the putative ΔHI0938 clones. For each DNA, the primers sets were used in the following order: PC1/PC2, P1/K2, and K1/P4. The positive and negative controls worked as expected: with 1219 bp in pLKR (lane 1) and 593 bp in KW20 DNA (lane 4); 2121 bp in pLKR (lane **2**), and no band for KW20 (lane **5**); 2202 bp for pLKR (3) and no band for KW20 (lane **6**). PCR for clone numbers P3-1, P3-2 (RR938), P4-1, and P4-2 worked with some primer sets, but not with others. Overall, however, the results indicate that all four of the above strains have HI0938 replaced by KAN. The clearest signals are from RR938, which had the expected amplicons for the three different PCRs, and from P4-1, which had the expected 1219 bp band using PC1 and PC2 primers. Although the PC1/PC2 primer set for P3-1 and P4-1 (but for P4-1 there was a very faint band of the correct size, lane **16**), did not yield detectable signals, PCRs for the other two primer sets did work, suggesting that KAN replaced the HI0938 gene, but have some sequence change in the LEFT region (e.g. similar to RR1172) which prevents PC1 annealing. Following these results, all of the above clones were tested for transformation defects as compared to the wild type strain, KW20.

Appendix B: Sxy and CRE Protein Sequence Analyses.

GENE	HI#	SignalP v 2.0 HMM		Cleavage Site					
		SP?	*SP Prob*	*NN max C*	*NN max Y*	*NN max S*	*NN mean S*	*HMM site&prob*	
comN	O938	SP	0.555	0.333 (56)	0.283 (25)	0.924 (19)	0.614 (1-24)	24	0.305
comO	O939	SP	0.988	0.163 (36)	0.314 (34)	0.974 (21)	0.827 (1-33)	33	0.62
comP	O940	SP	0.867	0.299 (23)	0.426 (23)	0.930 (15)	0.750 (1-22)	22	0.441
comQ	O941	SP	0.999	0.217 (36)	0.368 (36)	0.981 (20)	0.795 (1-35)	34	0.488
comA	O439	no	0	0.054 (60)	0.088 (10)	0.619 (2)	0.442 (1-9)		0
comB	O438	SP	0.931	0.116 (38)	0.272 (38)	0.980 (30)	0.630 (1-37)	40	0.558
comC	O437	no	0.001	0.155 (25)	0.290 (38)	0.908 (33)	0.431 (1-37)		0
comD	O436	SP	0.867	0.404 (63)	0.490 (19)	0.917 (4)	0.789 (1-18)	19	0.866
comE	O435	SP	0.999	0.619 (24)	0.668 (24)	0.932 (15)	0.790 (1-23)	24	0.999
comF	O434	no	0	0.093 (49)	0.106 (12)	0.580 (1)	0.423 (1-11)		0
comG	O433	no	0	0.241 (17)	0.158 (17)	0.227 (7)	0.137 (1-16)		0
comJ	O441	no	0	0.291 (11)	0.174 (11)	0.221 (25)	0.165 (1-10)		0
comE1	1008	SP	1.000	0.760 (22)	0.717 (22)	0.971 (10)	0.809 (1-21)	22	0.994
comM	1117	no	0.003	0.121 (17)	0.236 (14)	0.722 (5)	0.591 (1-13)	17	0.002
dprA	O985	no	0	0.102 (19)	0.118 (19)	0.465 (2)	0.249 (1-18)		0
pilA	O299	SP	0.727	0.205 (58)	0.218 (33)	0.958 (27)	0.615 (1-32)	32	0.325
pilB	O298	no	0.000	0.532 (15)	0.328 (15)	0.513 (5)	0.249 (1-14)		0
pilC	O297	no	0.000	0.083 (25)	0.090 (11)	0.453 (1)	0.362 (1-10)		0
pilD	O296	no	0.413	0.093 (35)	0.145 (19)	0.914 (10)	0.778 (1-18)	23	0.188
radC	O952	no	0.000	0.468 (70)	0.273 (70)	0.844 (1)	0.194 (1-69)		0
rec2	O061	no	0.401	0.159 (16)	0.220 (70)	0.966 (51)	0.819 (1-69)	16	0.257
ssb	O250	no	0	0.193 (30)	0.069 (3)	0.359 (1)	0.323 (1-2)		0
hyp	O365	no	0	0.153 (21)	0.072 (9)	0.411 (4)	0.293 (1-8)		0
hyp	O659	no	0	0.115 (64)	0.081 (12)	0.610 (1)	0.235 (1-11)		0
hyp	O660	no	0	0.052 (54)	0.048 (13)	0.159 (37)	0.106 (1-12)		0
hyp	1162	no	0	0.169 (10)	0.170 (10)	0.411 (30)	0.411 (1-9)		0
hyp	1182	SP	1	0.860 (19)	0.830 (19)	0.982 (11)	0.936 (1-18)	18	0.95
hyp	1183	no	0	0.085 (52)	0.109 (12)	0.289 (1)	0.197 (1-11)		0
sxy	O601	no	0	0.113 (66)	0.051 (66)	0.311 (64)	0.089 (1-65)		0

Appendix B: Sxy and CRE Protein Sequence Analyses.

SignalP Legend
NN=Neural Network
HMM= hidden Markov Model, pos. for max cleavage probability
SP Prob.=Signal Peptide probability (cut-off=0.5)

max C (cutoff=0.50) = cleavage site residue
max Y (cut off =0.32) = max Y score prob and position
max S (cutoff=0.90) = residue is part of SP or not
mean S (cutoff=0.44)

TMHMM v2.0 legend
TMH (transmembrane helix) cut-off = 0.5

Appendix B: Sxy and CRE Protein Sequence Analyses.

GENE	HI#	TMHMM v2.0 TMH?	PSORT localization	PSORT-B C	IM	P	OM	EX	Analysis:
comN	0938	1 (7-28)	OM 0.881, PS 0.270	7.31	0.67	0.98	0.54	0.51	Unk. (C)
comO	0939	1 (10-32)	IM 0.359	2	2	2	2	2	Unk. (1 TM)
comP	0940	1 (7-26)	IM 0.482	2	2	2	2	2	Unk. (1 TM)
comQ	0941	1 (13-35)	PS 0.789, OM 0.278	2	2	2	2	2	Unk. (1 TM)
comA	0439	no	cyto 0.315	7.31	0.67	0.98	0.54	0.51	Unk. (C)
comB	0438	1 (20-42)	IM 0.472	2	2	2	2	2	Unk. (1 TM)
comC	0437	1 (20-40)	IM 0.287	2	2	2	2	2	Unk. (1 TM)
comD	0436	no	PS 0.937, OM 0.296	7.31	0.67	0.98	0.54	0.51	Unk. (C)
comE	0435	1 (5-24)	PS 0.825, OM 0.198	0	0	0	10	0	OM
comF	0434	no	cyto 0.391	2	2	2	2	2	Unk. (C)
comG	0433	no	cyto 0.150	7.31	0.67	0.98	0.54	0.51	Unk. (C)
comJ	0441	no	IM 0.083	7.31	0.67	0.98	0.54	0.51	Unk. (C)
comE1	1008	no	PS 0.940, OM 0.334	4.14	1.46	2.13	1.17	1.1	Unk. (SP/C)
comM	1117	no	cyto 0.013	2	2	2	2	2	Unknown
dprA	0985	no	cyto 0.260	2	2	2	2	2	Unk. (Unk)
pilA	0299	1 (12-34)	OM 0.493, PS 0.303	2	2	2	2	2	Unk. (Unk)
pilB	0298	no	cyto 0.163	8.44	0.1	1.27	0.1	0.1	Cytoplasmic
pilC	0297	3 (162-184)	IM 0.601	0.18	9.02	0.1	0.35	0.35	IM
pilD	0296	6 (4-26)	IM 0.286	0	9.99	0	0	0	IM
radC	0952	no	cyto 0.156	7.31	0.67	0.98	0.54	0.51	Unk. (C)
rec2	0061	11 (5-22)	IM 0.667	0.18	9.02	0.1	0.35	0.35	IM
ssb	0250	no	cyto 0.078	2	2	2	2	2	Unk. (Unk)
hyp	0365	no	cyto 0.272	7.31	0.67	0.98	0.54	0.51	Unk. (C)
hyp	0659	no	cyto 0.455	2	2	2	2	2	Unk. (Unk.)
hyp	0660	no	cyto 0.371	7.31	0.67	0.98	0.54	0.51	Unk. (C)
hyp	1162	no	cyto 0.579	8.96	0.51	0.26	0.01	0.26	Cyto.
hyp	1182	no	PS 0.927, OM 0.190	4.14	1.46	2.13	1.17	1.1	Unk. (SP/C)
hyp	1183	no	cyto 0.380	2	2	2	2	2	Unk. (Unk)
sxy	0601	no	cyto 0.243	7.31	0.67	0.98	0.54	0.51	Unk. (C)

Appendix B: Sxy and CRE Protein Sequence Analyses.

PSORT legend	PSORT-B legend
IM=inner membrane	Unk.=unknown
OM=outer membrane	TM=has a Transmembrane Helix
PS=periplasmic space	SP=has Signal Peptide
cyto=cytoplasm	C=cytoplasmic
not shown='not clear'	P=found in Periplasm
	IM=found in Inner Membrane
	OM=found in Outer Membrane
	EX=Extracellular
	SP/C=predicted to have SP and be cyto.
	P/C=predicted in both Peri. and Cyto.