UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

FUNCTIONAL GENOMIC ANALYSIS OF *HAEMOPHILUS INFLUENZAE* AND APPLICATION TO THE STUDY OF COMPETENCE AND TRANSFORMATION

A Dissertation

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

Doctor of Philosophy

By

TIMOTHY M. VANWAGONER
Norman, Oklahoma
2004

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/761

UMI Number: 3138522

UMI®

UMI Microform 3138522

Copyright 2003 by ProQuest Information and Learning Company.

All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
PO Box 1346
Ann Arbor, MI 48106-1346

FUNCTIONAL GENOMIC ANALYSIS OF *HAEMOPHILUS INFLUENZAE* AND APPLICATION TO THE STUDY OF COMPETENCE AND TRANSFORMATION

A Dissertation APPROVED FOR THE
DEPARTMENT OF BOTANY AND MICROBIOLOGY

BY

_____
Jimmy D. Ballard, Ph.D., Chair

_____
John S. Downard, Ph.D.

_____
David S. Durica, Ph.D.

_____
William Ortiz-Leduc, Ph.D.

_____
Terrence L. Stull, M.D.

© Copyright by TIMOTHY M. VANWAGONER 2004
All Rights Reserved

ACKNOWLEDGEMENTS

I have had the benefit of interacting with so many knowledgeable individuals during my studies that I could never begin to thank them all. However, I am eternally grateful to three people who acted as mentors to me as I pursued my degree. First, I must thank David McCarthy, who was my first advisor. His guidance was critical to my learning to think big. Without his mentorship, my interest in functional genomics may never have developed. Second, I am so appreciative of the mentorship offered to me by Terrence Stull. He accepted me into his laboratory to finish my studies at a time when circumstances had led me to be very disenchanted with my career choice. His unyielding support was vital to my continued development as a scientist and this work could never have been completed without him. Last, but not least, I am forever indebted to Paul Whitby for his friendship, support and mentoring. It has been wonderful having such a generous and conscientious fellow looking out for me and my wife. Thanks Terry and Paul for helping me rediscover my love of science.

I would also like to extend my thanks to my dissertation committee members, past and present, for all of their assistance. I must also thank the Department of Botany and Microbiology for its continued support over the years. I learned, much to my surprise, how much I truly enjoy teaching and the department gave me plenty of opportunities to hone those skills

I want to thank several others that I have interacted with during my studies. My original work with Dave was towards the development of helicase activity gels and Sanjay Shukla was patient with me when I was trying to learn a very difficult protocol. I

also want to thank those with whom I have had the pleasure of interacting with since joining Terry's group, in particular Drs. Danny Morton and Tom Seale.

Most importantly, I want to acknowledge my wonderful wife Aimee for her undying support and encouragement during my studies. Since I started, she has managed to finish two master's degrees and has been kind enough not to remind me of that fact too often. She has never been an impediment to my work and never complained about the long hours I often kept or when we have had to skip family holidays because I needed to perform some experiment. This dissertation is lovingly dedicated to her.

TABLE OF CONTENTS

List of Tables viii

List of Figures ix

Abstract xi

Part I – Functional Genomics

Chapter 1. General introduction: Functional genomics 1

Gene disruption strategies 3

Gene expression analysis 8

Analysis of protein expression and function 16

In silico analysis 19

Chapter 2. Transposon mutagenesis of *Haemophilus influenzae* 21

Abstract 21

Introduction 21

Materials and methods 25

Results 46

Discussion 85

References (Part I) 95

Part II – Competence and transformation

Chapter 3. General introduction: Competence and transformation 105

DNA binding 109

DNA uptake and translocation 109

Recombination of DNA 119

Regulation of competence development 121

Chapter 4. Competence and transformation studies in *Haemophilus influenzae* 130

Abstract 130

Introduction 131

Materials and methods 133

Results 150

Discussion 172

References (Part II) 182

Appendices

Appendix A – Mapped Tn7 insertions (pASC13 libraries) 198

Appendix B – Mapped Tn7 insertions (pASC15 libraries) 201

Appendix C – Mapped Tn5 insertions (pASC15 libraries) 204

Appendix D – Mapped Tn5 insertions (pASC18, pre-mutagenesis minimalization) 210

Appendix E – Mapped Tn5 insertions (pASC18, post-mutagenesis minimalization) 229

Appendix F – Mapped Tn5 insertions (pASC18MIN libraries) 238

LIST OF TABLES

Table		
2.1	Strains and plasmids used in this work	26
2.2	Nucleotide sequences of primers and adapters used in this work	27
2.3	Frequency of nucleotides at positions surrounding the Tn5 insertion sites in Rd KW20 DNA	61
2.4	Distribution of Tn5 insertions into Rd KW20 CDS in relation to GC content	64
3.1	Genes implicated in competence and transformation in *H. influenzae*	107
4.1	Strains and plasmids used in this work	134
4.2	Nucleotide sequences of primers used in this work	136
4.3	Pasteurellaceae genomic sequences examined in this study	155
4.4	Examination of transformation efficiency, DNA binding and uptake for wild-type and mutant strains of *H. influenzae* Rd KW20	166
4.5	Examination of gene transcription in wild-type and mutant strains of *H. influenzae* Rd KW20 after 60 minutes incubation in MIV media	167
4.6	Examination of *comA* and HI0366 transcription in Rd KW20 and TMV15	168

LIST OF FIGURES

Figure		
1.1	Signature-tagged mutagenesis (STM)	5
1.2	GAMBIT	7
1.3	Gene expression studies using microarrays	10
1.4	Serial analysis of gene expression (SAGE)	13
1.5	Differential Display RT-PCR	15
1.6	Proteomic analysis by two-dimensional electrophoresis	18
2.1	Creation of the *Haemophilus influenzae* Rd KW20 insertion libraries	23
2.2	Development of the pASC family of minimal cloning vectors	30
2.3	Single primer mapping of insertion sites	41
2.4	Dual primer convoluted sequence mapping	43
2.5	Distribution of Tn7 insertions in pASC vectors	52
2.6	Distribution of Tn5 insertions in pASC vectors	53
2.7	Average GC content surrounding the Tn5 insertion sites in *H. influenzae* Rd KW20 genomic and vector DNA sequences	59
2.8	Average trinucleotide GC content surrounding the Tn5 insertion sites in *H. influenzae* Rd KW20 genomic and vector DNA sequences	67
2.9	Average A-philicity surrounding the Tn5 insertion sites in *H. influenzae* Rd KW20 genomic and vector DNA sequences	69
2.10	Average protein-induced deformability surrounding the Tn5 insertion Sites in *H. influenzae* Rd KW20 genomic and vector DNA sequences	71
2.11	Average B-DNA twist surrounding the Tn5 insertion sites in *H. influenzae* Rd KW20 genomic and vector DNA sequences	73
2.12	Average bendability surrounding the Tn5 insertion sites in *H. influenzae* Rd KW20 genomic and vector DNA sequences	75

LIST OF FIGURES (CONTINUED)

2.13	Bendability and GC content plots for several Rd KW20 low G+C CDSs with recovered Tn5 insertions	77
2.14	Bendability and GC content plots for several Rd KW20 CDSs with multiple recovered Tn5 insertions	80
2.15	Bendability and GC content plot of an Rd KW20 rRNA operon	84
3.1	Proposed model of the DNA uptake mechanism in *H. influenzae*	113
4.1	Putative CRE regions in *H. influenzae* Rd KW20	132
4.2	Screening transposon libraries for transformation mutants	152
4.3	Organization and conservation of the CRE0364 region in the family Pasteurellaceae	156
4.4	Organization and conservation of the CRE0937 region in the family Pasteurellaceae	157
4.5	Organization and conservation of the CRE1181 region in the family Pasteurellaceae	158
4.6	Expression profile of *tfoX, rec-2,* and *comA* during competence development	161
4.7	Expression profile of genes contiguous with CRE0364 during competence development	162
4.8	Expression profile of genes contiguous with CRE0937 during competence development	163
4.9	Expression profile of genes contiguous with CRE1181 during competence development	164
4.10	Locations of Tn5 insertions in HI1159m, *hemH* and HI1161	171
4.11	Comparison of the HI0366 locus in *H. influenzae* Rd KW20 with the *pilF* locus of *P. aeruginosa* PA01	175

ABSTRACT

The publication of the complete genomic sequence of *Haemophilus influenzae* Rd KW20 in 1995 was a truly monumental event in molecular biology. For the first time, all of the potential genes of an independent-living organism were known and awaiting functional characterization. This event required the development of fundamentally different methodologies to elucidate gene functions, with systematic global approaches becoming much more feasible. This study describes the development of a transposon-based mutagenesis strategy to facilitate a high-throughput functional analysis of the *H. influenzae* genome. Mutants created using this strategy were screened in a highly-parallel assay to identify genes mediating transformation in this organism. Additionally, analysis of the transposon insertion sites generated during this study identified a previously unrecognized Tn5 insertion bias.

During the progression of this study, hundreds of additional bacterial genomes were sequenced, techniques have evolved, and novel approaches were developed to aid the field of functional and comparative genomics. Some of these techniques were used in this study to identify three novel competence-regulated operons in *H. influenzae*. The techniques included the use of advanced computer programs and algorithms to assist in predicting protein functions and to facilitate a comparative genomic analysis of *H. influenzae* with other species of the Pasteurellaceae family. Quantitative PCR was employed to examine the expression of putative transformation-related genes. Finally, PCR-mediated mutagenesis was used in a directed approach to generate mutations in the newly discovered competence-regulated operons to assess their involvement in uptake and transformation of exogenous DNA.

PART I – FUNCTIONAL GENOMICS

CHAPTER ONE

General introduction: Functional Genomics

The publication of the first complete genomic sequence, *Haemophilus influenzae* Rd KW20 in 1995 (27), opened a new era in molecular biology by giving researchers access to the entire genetic complement of a living organism. Since that time, sequences of more than one hundred additional eubacterial, archaebacterial and eukaryotic genomes, including the human genome, have been completed and are available in the public domain. Countless other sequencing projects are in progress and data is rapidly accumulating. The preeminent issue currently facing molecular biologists is how to use the large amount of data to assist in determining the functions of the newly identified genes. Accomplishing this task will not only help in understanding the basic processes of life but also fighting human disease, combating microbial infections and food spoilage, and genetically engineering microbes for industrial processes such as bioremediation. The magnitude of the task is daunting. For example, the original analysis of the *H. influenzae* sequencing data indicated the genome consisted of 1743 predicted coding regions. General role assignments were postulated for 1007 of the predicted protein coding sequences (CDSs) but the exact role of many of these remain unknown. No role assignment was given for 736 (42%) of the annotated CDSs (27). The genomic sequence of the yeast *Saccharomyces cerevisiae* consists of approximately 13-Mbp and 6000 predicted CDSs (67). By mid 1997, analysis of the yeast genome revealed that only 46% of predicted proteins had been biologically characterized and that the function of an

additional 22% could be inferred by homology to other experimentally characterized proteins, thus leaving more than 30% of the yeast genes uncharacterized (109).

The burgeoning field of functional genomics is devoted to determining the functions of the uncharacterized CDSs and to determine the essentiality and expression patterns of these genes. A large array of techniques exists and more are under development to advance these goals. In general, these analyses fall into four categories:

1. gene disruption analysis
2. transcriptomics
3. proteomics
4. *in silico* analysis.

In the first, predicted genes are disrupted by either deletion or insertion mutagenesis and the phenotypic result of the loss of the gene product is used as the first step in determining its potential role. The disruption analysis can also be used to determine the essential nature of the gene (either complete or conditional lethality) which can assist in defining potential drug or vaccine targets. In the second, gene expression patterns are studied by determining conditional expression of transcripts to infer a possible role for the gene. Similarly, the field of proteomics includes methodologies to examine the conditional expression of proteins. Additionally, proteomic technologies have been developed to elucidate protein functions through localization studies or protein-protein interactions. Finally, computational analysis is used to compare potential gene products to known proteins to assist in assigning function, to recognize patterns such as export signals that might help determine cellular localization, and to compare genomic and proteomic sequences for examination of evolutionary relationships and horizontal gene

transfers. Examples from several of these categories have been used in this work to study functional genomics in *Haemophilus influenzae.*

Gene disruption strategies. Since the development of basic molecular tools, gene disruption has been the most accessible technique available for genomic analysis. The construction and analysis of mutants affords the flexibility to be employed at both small and large scales. The availability of complete genomic sequences allows the process to be performed at a higher pace by simplifying the mapping of random insertions or targeting of gene deletions. Transposon insertion mutagenesis (TIM) has figured prominently in functional genomic strategies for several reasons. Transposable elements, with selectable markers for easy isolation of mutants, can be easily used to generate disruptions in a gene, and any resulting phenotypic changes may help identify the role of the gene product. Finally, the transposon itself is an island of unique sequence that facilitates the determination of the location of the insertion site by sequencing or PCR-based mapping. Traditional *in vivo* transposon mutagenesis has been limited by the requirement either to clone the genomic sequence into plasmids, with subsequent mutagenesis within *Escherichia coli*, or to find a transposon system compatible with the host. For the latter to work, the transposase must be expressed and functional in the host. Such a system also must be engineered to separate the transposase gene from the insertion element or the insertion may be unstable. These limitations are largely alleviated by the development of *in vitro* transposon systems utilizing purified transposases (46). These transposases are engineered to remove most of the native sequence insertion biases or to develop hyperactive transposition mutants for greater yields. In addition, the transposable elements are designed with unique priming sites at

their termini and other features, including variable selectable markers and inducible promoters. The most prominently used *in vitro* systems are the commercially available Tn5 (Epicentre) (39), Tn7 (New England Biolabs) (12, 95) and Ty1 (AppliedBiosystems) (32) kits and purified Himar-1 *(mariner)* transposase (55). Furthermore, the development of purified transposomes, composed of transposase-transposon complexes, overcomes the limitations of traditional *in vivo* transposition by allowing electroporation of the complexes directly into the organism (37, 54).

The availability of transposon tools and complete genomic sequences has led to numerous publications detailing their use in small-scale studies in which mutants with desired phenotypes are isolated and the site of insertion subsequently mapped. Of greater interest to this work are global approaches using transposon mutagenesis. Previously, these studies have fallen into three related, yet distinct lines: signature-tagged mutagenesis (STM), footprinting strategies, and arrayed mutant libraries.

Transposon-based STM is designed as a high-throughput method to identify conditionally essential genes, especially virulence determinants (Figure 1.1). A set of transposons containing unique nucleotide regions serve as virtual barcodes to identify individual insertions. The mutants are then pooled into groups and selective pressure is applied. The surviving organisms are collected and the tags are amplified by PCR and hybridized to an array composed of oligonucleotides complementary to the tag sequences. The absence or dramatic decrease of a tag indicates a mutation detrimental to growth under the given conditions, and the site of the mutation can be mapped to identify the gene involved. This highly parallel strategy was first employed to determine virulence

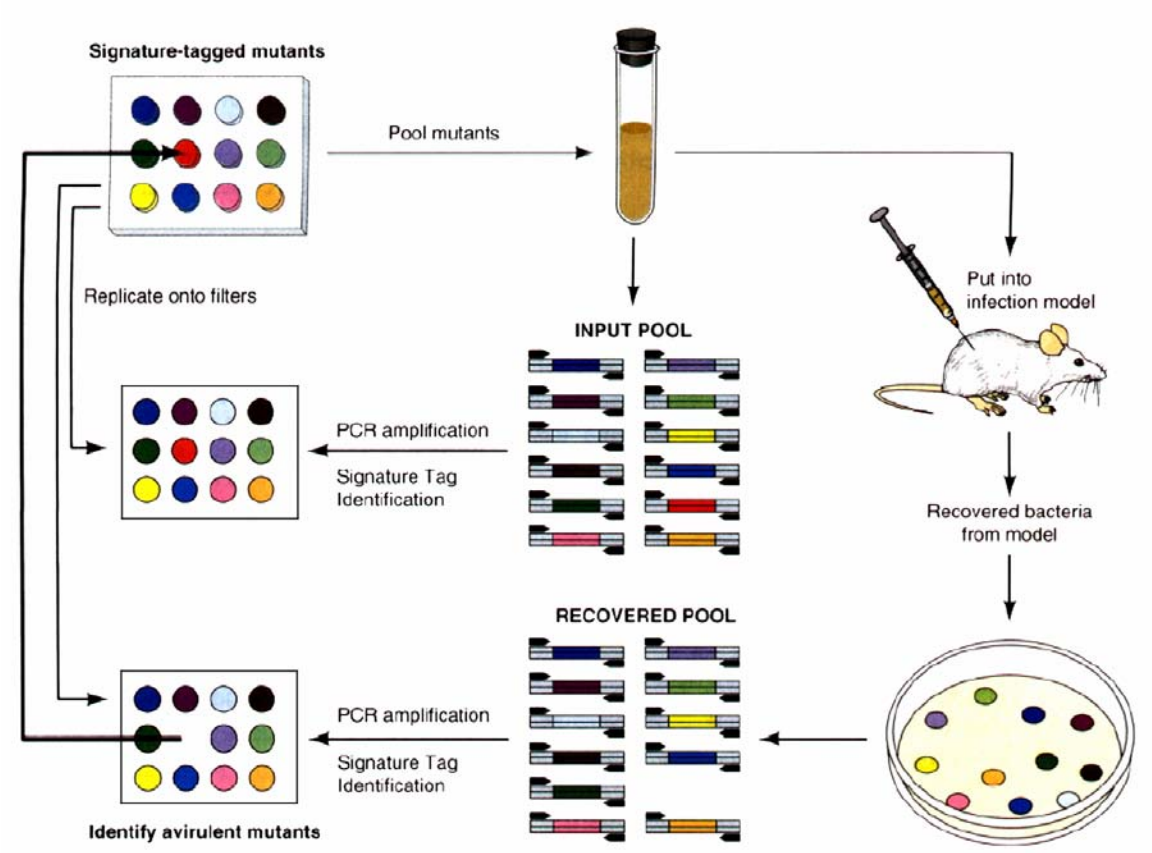


Figure 1.1 – Signature-tagged mutagenesis (STM). Example of a typical STM screen. A pool of signature-tagged transposon mutants is arrayed, and DNA colony blots of the individual mutants are generated by replica plating. The mutants are pooled for use as an inoculum. Genomic DNA is prepared from a sample of the inoculum and signature tags from the pool are amplified and radiolabeled by PCR in order to characterize the input pool. The inoculum is then injected into a suitable animal model. Following the course of the disease, bacteria are harvested from the animal and signature tags amplified to create the output pool. Comparision of the input hybridization blots with the output blots identifies mutants with attenuated virulence. Figure from Shea and Holden (89)

factors of *Salmonella typhimurium* in the murine model of typhoid fever (48), and has since been successfully employed to study virulence in a large number of pathogens, including *H. influenzae* (49) and other members of the Pasteurellaceae (29, 30).

A second transposon-based approach, genetic footprinting, was first used to help identify gene functions in *S. cerevisiae* (92, 93) and has been employed in *E. coli* (33, 44). Saturation mutagenesis is used to create a massive pool of insertion mutants, and a sample is taken to represent the population at time zero. The population is then exposed to a selective condition and samples taken following outgrowth. The samples are compared by analyzing PCR products generated using a transposon-specific primer and varying genome-specific primers. When analyzed by gel electrophoresis, this results in a ladder of bands corresponding to transposon insertion sites. Comparison of banding patterns from the initial culture to samples taken following selective pressure can identify sites of insertion that resulted in a loss of fitness in the mutant, identifying genes likely to be important under those conditions. A variant of this technique, genomic analysis and mapping by *in vitro* transposition (GAMBIT), utilizes *in vitro* transposition of defined PCR products (~10 kb) and natural genetic transformation to transfer the insertions into the host (Figure 1.2) (3). Genes for which no, or few, insertions can be recovered are putatively defined as essential to the organism. This approach has been employed on a limited scale in several studies and on a large scale recently in *H. influenzae* (4).

A third transposon-based approach is the systematic creation of defined insertion mutants that can be examined for disruption phenotypes. Genomic DNA is cloned into a suitable vector and mutagenized *in vitro* or in *E. coli.* The transposon

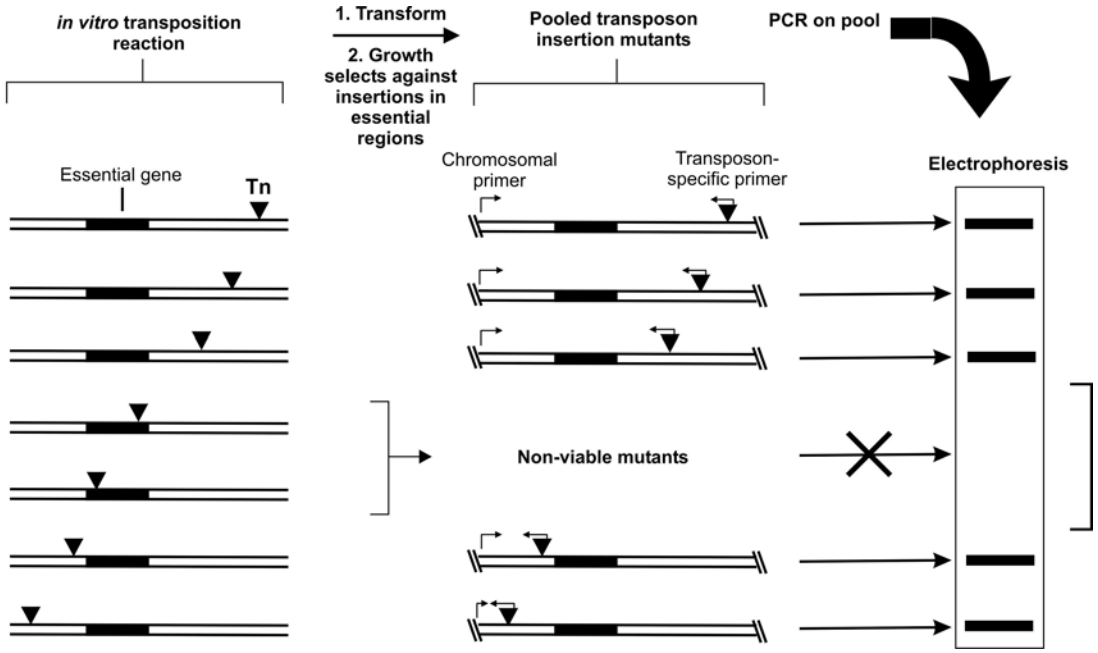


Figure 1.2 – GAMBIT. An example of genomic footprinting. Defined pools of target DNA are mutagenized by *in vitro* transposon mutagenesis and introduced into bacteria by transformation. Mutants harboring insertions in essential genes are lost from the pool following outgrowth. PCR using a transposon-specific primer and specific chromosomal primers is performed and the products resolved by gel electrophoresis. PCR products corresponding to insertions in non-essential genes will appear as band on the gel. Non-permissive mutations will result in gaps in the electrophoretic-profile. Figure adapted from Akerley *et al.* (3)

insertion sites are mapped, and the mutation is transferred back to the organism by homologous recombination. The result is a comprehensive collection of insertion mutants in non-essential genes that can be used for functional annotation of the genome. This strategy has been predominantly employed in eukaryotic studies (17, 42, 82) and is a focus of this work.

Another common means of gene disruption utilized in genomic analysis is the directed deletion of specific genes. This has traditionally been accomplished by cloning a fragment of the genome carrying the gene followed by restriction digestion to remove the gene and insertion of a selectable marker in its place. PCR-mediated deletion improves the efficiency of this approach (100). Gene deletion in the post-genome era is an important tool for functional analysis; however, costs have limited its use in large-scale approaches. Like transposon mutagenesis, deletion analysis is also amenable to use in signature-tagged mutagenesis. The most extensive deletion analysis study was in *S. cerevisiae* (91, 110) and resulted in the deletion of 96.5% of the annotated genes in that organism (34). Since each deletion mutant contains a unique sequence tag, the ability to perform massive parallel analysis on the library could facilitate a more rapid functional annotation of the yeast genomic sequence.

Gene expression analysis. Examination of expression patterns has played an important role in the study of genes. Prior to the publication of whole genome sequences, these studies predominantly used reverse-transcriptase (RT) mediated polymerase chain reaction (RT-PCR) analysis of transcript production or the use of fusion-reporter constructs in either a directed (21, 79) or blind approach (5). True global analysis of expression patterns was limited to proteomic approaches with two-dimensional gel

electrophoresis, and even this method has been greatly enhanced by the availability of genomic data. Several new, high-throughput technologies have been developed, including DNA microarrays, differential display, and Serial Analysis of Gene Expression (SAGE), to take advantage of complete genomic information.

Of these technologies, DNA microarrays have emerged as a powerful means of exploring transcript production patterns under varying environmental conditions (Figure 1.3) (18, 60). Briefly, either synthetic oligonucleotides or PCR-generated fragments complementary to each gene are immobilized in an ordered pattern on a glass or nylon membrane matrix. RNA is harvested from cells and fluorescently labeled either directly or following conversion into DNA (cDNA). The labeled molecules are then hybridized to the array and the signal intensities corresponding to each spot are measured. Following normalization, comparison of spot intensities between experimental conditions is used to determine which genes are up- or down-regulated under those conditions. The technique was first deployed in studies of *Arabidopsis thaliana* (86) and *S. cerevisiae* (24) but it has found increased use in the study of bacterial gene expression, including in *E. coli* (80, 97), *Mycobacterium tuberculosis* (14, 108), and *Pasteurella multocida* (73, 74). In addition, DNA microarrays have been used to study regulation associated with development of transformation competence in *Streptococcus pneumoniae* (75, 81) and *Bacillus subtilis* (15, 50, 52, 70, 71).

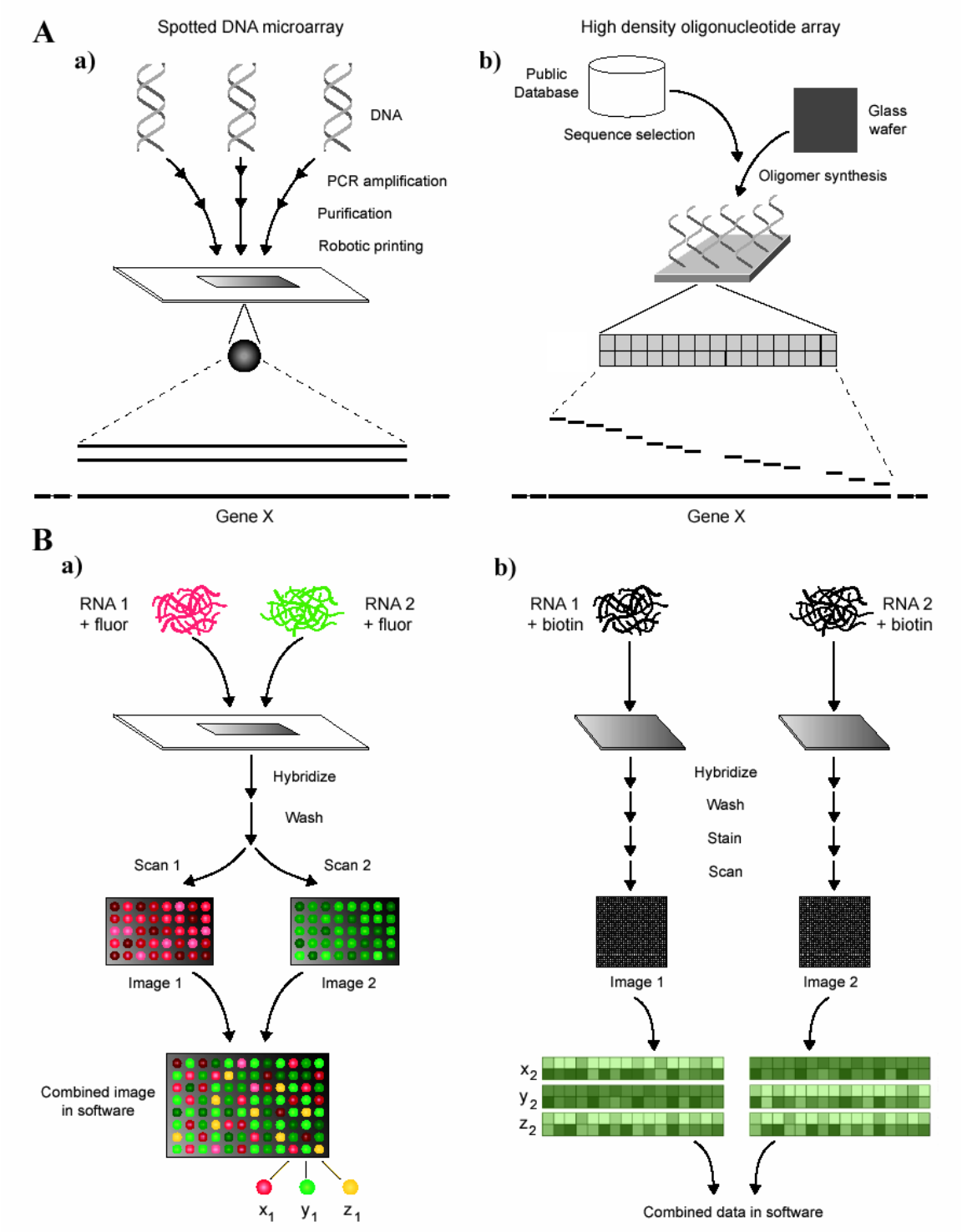
DNA microarrays are more suited for bacterial studies than its counterpart SAGE (Figure 1.4) (103). In SAGE, cDNA is synthesized using a biotinylated oligo(dT) primer and digested using a restriction endonuclease (anchoring enzyme, AE) with a 4-bp recognition site to increase the likelihood of cleaving each transcript at least once. The

Figure 1.3 – Gene expression studies using microarrays.

A. Preparation of microarray chips. (a) Lower density, spotted array chips are prepared by robotic placement of amplified DNAs or synthesized oligonucleotides onto a glass slide. Each spot can correspond to large or small fragments of a gene. **(b)** High-density arrays are prepared by *in situ* synthesis of oligonucleotides on a chip. Each gene is represented by numerous oligonucleotide probes.

B. Assay of gene expression. (a) Using spotted arrays, mRNA from test and reference samples are labeled either directly or indirectly (via cDNA synthesis) using different colored fluorophores. The mixtures are hybridized to the glass slides and scanned to detect levels of each fluorophore at each spot. Colored dots x, y and z correspond to RNA levels of hypothetical genes. Increased levels in sample 1 are represented by x_1, increased levels in sample 2 by y_1 and similar levels in both samples by z_1. **(b)** Using high-density arrays, mRNA is biotin labeled in a linear amplification process. Each sample is hybridized to separate chips and stained with an avidin-conjugated fluorophore. Sets of oligonucleotides corresponding to hypothetical genes are shown below the chip diagrams. Increased levels of intensity (corresponding to increased expression) in sample 1 are represented by x_2, increased levels in sample 2 by y_2, and similar levels in both samples by z_2.

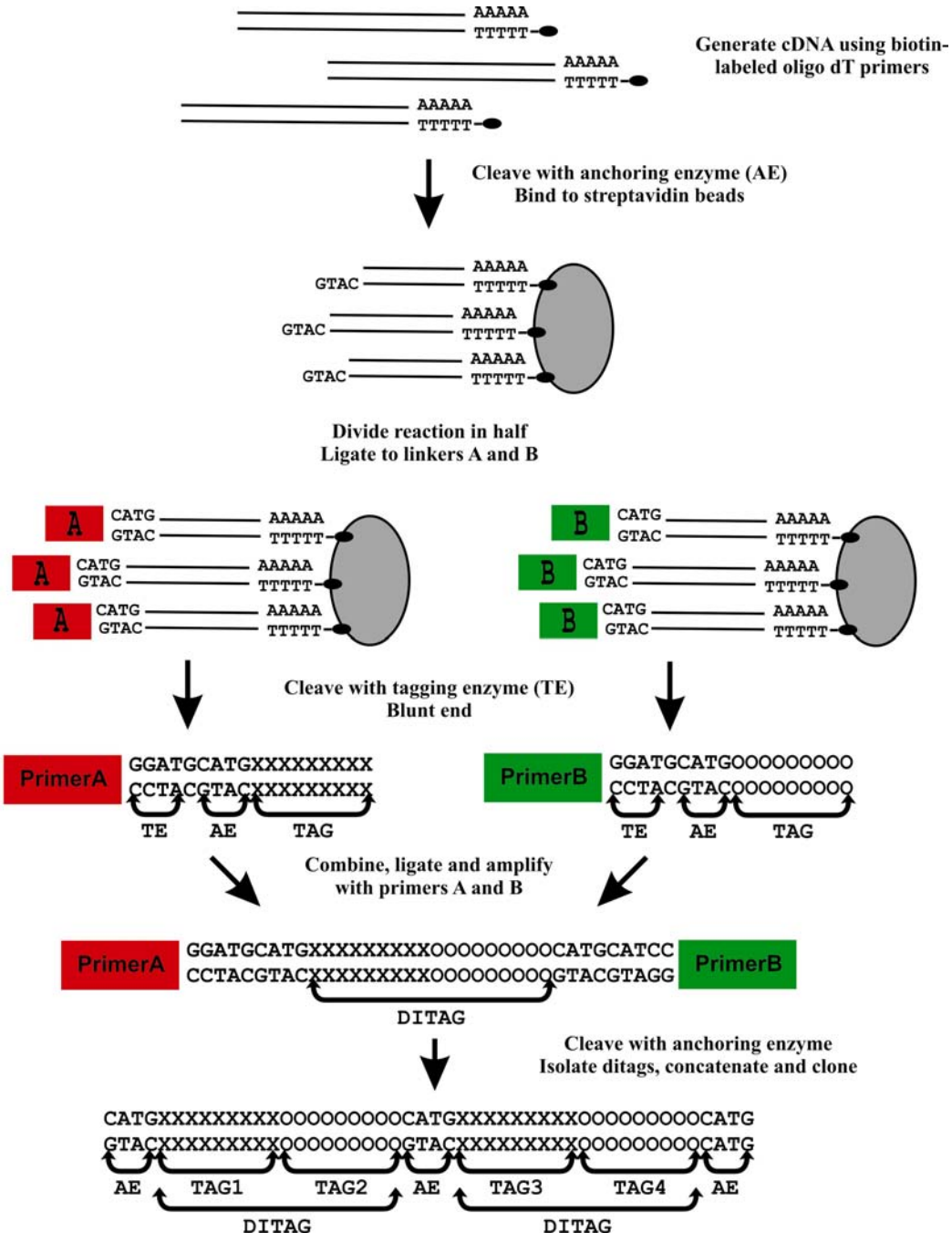
Figure adapted from Harrington *et al.* (45)


A
Spotted DNA microarray
a)
DNA
PCR amplification
Purification
Robotic printing
Gene X
High density oligonucleotide array
b)
Public Database
Sequence selection
Glass wafer
Oligomer synthesis
Gene X
B
a)
RNA 1 + fluor
RNA 2 + fluor
Hybridize
Wash
Scan 1
Scan 2
Image 1
Image 2
Combined image in software
x1
y1
z1
b)
RNA 1 + biotin
RNA 2 + biotin
Hybridize
Wash
Stain
Scan
Image 1
Image 2
x2
y2
z2
Combined data in software

cDNA is isolated by binding to streptavidin beads and divided into two populations. Each half of the population is ligated to one of two linkers that consist of sequence complementary to the AE site, a type IIS restriction site (tagging enzyme, TE) and a unique primer binding site. The TE site is situated so that the enzyme cleaves at a defined distance from the recognition site to leave a short fragment of DNA that includes a 9 bp tag derived from the original transcript. After TE cleavage, the two populations are mixed, blunt-end ligated and amplified with the two primers. These amplicons are then recleaved with the AE to form a fragment that contains terminal sticky ends and a ditag that represents two separate transcripts. These fragments are then concatenated and sequenced, with the AE sites acting as punctuation marks, to determine the relative occurrences of transcripts in the original population. In theory, the abundance of tags appearing in the sequences allows SAGE to provide quantitative information on gene expression. A variation on the original method, LongSAGE, generates longer tags (21bp) for increased ability to assign tags to their correct transcript (84). While SAGE has become extremely popular in transcriptome analysis in humans and other eukaryotes, the reliance on poly-adenylated transcripts limits its usefulness in prokaryotic studies.

A third common methodology for transcriptome analysis is differential display (DD) (Figure 1.5) and its related RNA fingerprinting techniques (57, 58, 63, 107). In short, RT-PCR is performed on the RNA sample using an oligo(dT) primer or a primer composed of arbitrary nucleotides linked to a defined sequence (anchor primer site). Second strand synthesis and amplification is carried out at low stringency with an arbitrary primer and the anchor primer. This is followed by additional rounds of amplification at high-stringency using the same primers. The reactions are resolved on a

Figure 1.4 – Serial analysis of gene expression (SAGE). cDNA is synthesized using biotin-labeled oligo dT primers, cleaved with an anchoring restriction enzyme and bound to streptavidin magnetic beads. The reaction is divided and ligated to either linker A or linker B. Each linker contains a unique primer binding site and a recognition site for a type II restriction enzyme (the tagging enzyme). The reactions are cleaved with the tagging enzyme, blunt-ended and the remainder of the cDNA fragments removed by magnetic separation. The reactions are combined, ligated together and amplified with primers specific to linker A and linker B. The PCR products are then cleaved with the anchoring enzyme, concatenated and cloned into an appropriate vector for DNA sequencing. The sequencing data consists of multiple ditags separated by the anchoring enzyme recognition sites. Each ditag is composed of two 9-bp tags derived from the original cDNA. If the locations of the anchoring enzyme sites are known for each gene, the gene corresponding to the tag can be identified. Relative abundance of the tags can be used to quantify changes in conditional gene expression. Figure adapted from Velculescu *et al.* (103)


AAAAA
TTTTT
Generate cDNA using biotin-labeled oligo dT primers
Cleave with anchoring enzyme (AE)
Bind to streptavidin beads
GTAC
Divide reaction in half
Ligate to linkers A and B
A
B
CATG
Cleave with tagging enzyme (TE)
Blunt end
PrimerA
PrimerB
GGATGCATGXXXXXXXXX
CCTACGTACXXXXXXXXX
GGATGCATGOOOOOOOOO
CCTACGTACOOOOOOOOO
TE
AE
TAG
Combine, ligate and amplify
with primers A and B
GGATGCATGXXXXXXXXXOOOOOOOOOCATGCATCC
CCTACGTACXXXXXXXXXOOOOOOOOOGTACGTAGG
DITAG
Cleave with anchoring enzyme
Isolate ditags, concatenate and clone
CATGXXXXXXXXXOOOOOOOOOCATGXXXXXXXXXOOOOOOOOOCATG
GTACXXXXXXXXXOOOOOOOOOGTACXXXXXXXXXOOOOOOOOOCATG
TAG1
TAG2
TAG3
TAG4

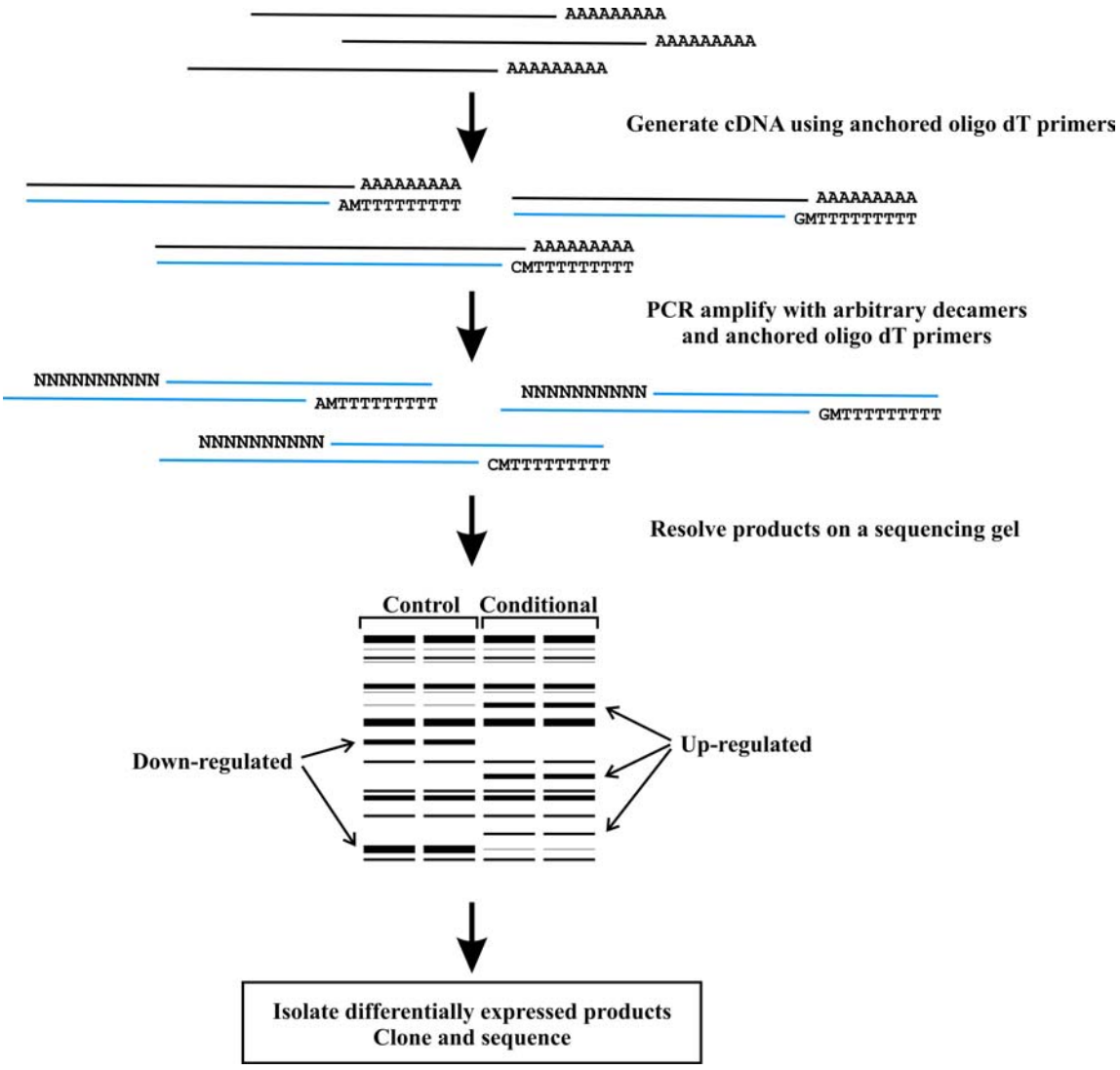


Figure 1.5 – Differential Display RT-PCR. RNA from cells grown under different conditions is divided into pools. cDNA is generated using different anchored oligo[dT] primers for each pool. The cDNA is then amplified using arbitrary decamers (or longer oligonucleotides) and the original anchored oligo[dT] primer used for that pool. Products are resolved by PAGE and differentially expressed bands are excised, cloned and sequenced to determine the gene from which they were derived.

sequencing gel resulting in a ladder of PCR fragments. When reactions from cultures grown under different conditions are run in parallel and compared, differential bands can indicate transcripts preferentially produced in response to the conditions. The differentially expressed bands are then excised, amplified with the same primers and sequenced to determine the corresponding gene. This technique also lends itself well to studies involving smaller numbers of genes, or to prokaryotic studies, using defined, gene-specific primers. Differential display has been used successfully in prokaryotic gene regulation studies, including in *Streptococcus mutans* (23, 43) and *H. influenzae* (96).

On a much smaller scale, quantitative RT-PCR (Q-PCR) has emerged as an important tool for analysis of expression of individual genes (35). In Q-PCR, reverse transcription is employed to generate cDNA from RNA followed by PCR with gene-specific primers and fluorescent probes or intercalating agents to monitor the increase in replicons following each PCR cycle. Quantification of RNA in the original sample is determined by comparison to known quantities of the RNA or to constitutively expressed RNAs. While Q-PCR has been predominately employed in the study of eukaryotic gene expression, it has been employed in prokaryotic research for validation of microarray expression data (47, 66, 78) and examination of individual gene expression (20, 65, 69, 83, 85).

Analysis of protein expression and function. The large-scale analysis of protein expression complements gene expression analysis. While production of mRNA is informative, it may not always correlate with actual protein production due to post-transcriptional regulatory events (40). Thus, proteomics would be the preferred methodology for characterizing gene functions in many situations. Unfortunately, the

methods available in proteomics are more difficult to perform and interpret than those in use for gene expression studies. The traditional tool for global analysis of protein expression has been two-dimensional (2-D) gel electrophoresis (Figure 1.6). In this method, proteins are first separated by isoelectric focusing and then by molecular mass using denaturing polyacrylamide gel electrophoresis (PAGE). The inherent problem of 2-D PAGE is that it requires a large quantity of protein and is difficult to reproduce without considerable expertise. Additionally, its dynamic range is somewhat limited and the methodology is biased towards soluble and abundant proteins (112). A new technology, 2D-fluorescence-difference gel electrophoresis (DIGE) is an improvement on traditional 2-D PAGE. Two protein samples from different conditions are labeled *in vitro* with different fluorescent dyes and run together in a single PAGE gel. Since both the control and experimental samples are run concurrently, DIGE is more reliable than its predecessor and the use of fluorescent-based detection makes it more sensitive for quantitation purposes (40). In both 2D-PAGE and DIGE, spots of interest are excised from the gel and subjected to proteolytic degradation. The resulting pool of peptides are then identified by mass spectrometry and compared to predicted proteins derived from genomic sequencing (2, 26). 2-D PAGE has been employed to define the proteome and to study differential expression in a number of bacterial species including *H. influenzae* (25, 56, 101).

Liquid chromatography (LC) mediated separation of proteins has become an alternative to electrophoretic separation methods. LC can be coupled with tandem mass spectrometry analysis (LC-MS/MS) to separate, fragment, and generate sequence data for proteins in the mixture (2, 112). While LC-MS/MS is more sensitive than gel

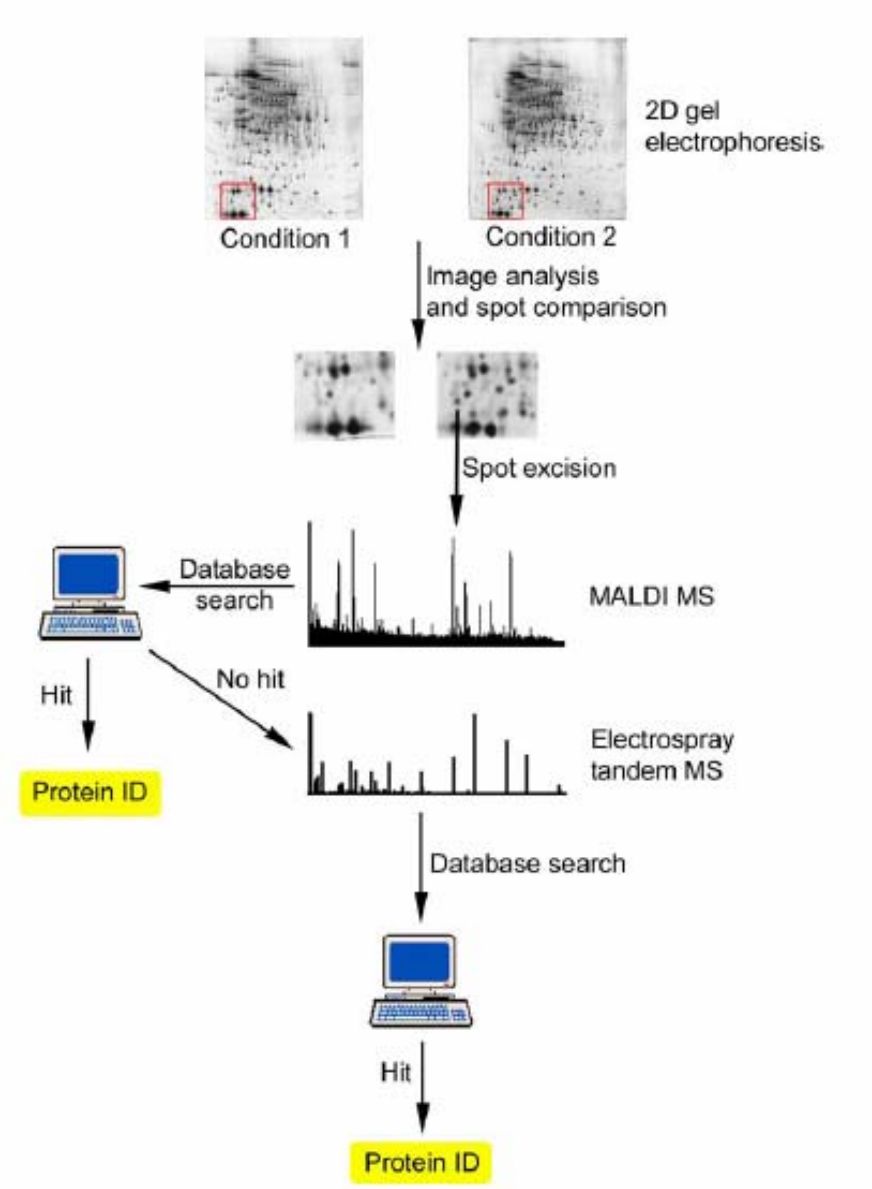


Figure 1.6 – Proteomic analysis by two-dimensional electrophoresis. Proteins isolated from cells under different conditions are separated by 2D electrophoresis. Spot patterns are compared for the different conditions. Novel spots, or spots of increased intensity, are cut from the gel and the protein (or peptide) identified by mass spectroscopy (MS), using either MALDI-TOF (matrix assisted laser desorption ionization, time-of-flight) or tandem MS. Figure from Betts (16).

electrophoresis methods and thus more capable of identification of low-abundance proteins, it is not suited for quantification studies. A variation of the method involves the addition of different isotope-coded affinity tags (ICAT) to proteins isolated from control and experimental samples. The samples are mixed and peptides separated by LC followed by tandem MS. The ratio of the signal intensities between the labeled peptides corresponds to the ratio of the abundance of the originating protein between the two samples (72). While ICAT technology has not been utilized for proteomic analysis in *H. influenzae*, LC-MS/MS has been employed in studies of the organism and shown to identify expressed proteins not isolated by 2D-electrophoresis (53).

In addition to expression proteomics, numerous methodologies are available or being developed to obtain data on the function of proteins. These include technologies such as two-hybrid studies, affinity tagging or protein arrays for dissecting protein-protein or protein-substrate interactions, epitope tagging for determination of protein localization and scanning linker mutagenesis for structure-function studies of individual proteins (46, 112). Of these, only two-hybrid systems have been extensively used in bacterial studies. In short, a protein or peptide of interest (bait) is fused to a DNA binding domain and other proteins of interest (prey) are fused to a transcription factor activation domain. The interaction between bait and prey brings these two domains into proximity and leads to transcription from a reporter construct. This method is amenable to both smaller-scale (51, 64) and global studies (77) of bacterial protein interactions.

In-silico analysis. The accumulation of genomic sequences, expression data, protein interaction data and elucidated protein structures, combined with increased computer power, has led to the increased use of computers to assist in characterizing the

function of unknown gene products. Among the first of these approaches was the BLAST algorithms and databases that allow for alignment of DNA and protein sequences across species (7). Another such method is the identification of possible functional domains or motifs of proteins by exhaustive comparison among proteins. The COG, PROSITE and pfam databases are examples of classification schemes that rely on primary structural analysis of proteins, and the SCOP database is an example of schemes that rely on tertiary structure (13, 68, 94, 99).

Many techniques have been developed to exploit the rapidly accumulating genomic data. These techniques include strategies for gene disruption, transcriptional and proteomic analyses and *in silico* predictions of protein functions. Determination of the functions of all the putative genes in each bacterial genome is likely to require a combination of methods. Chapter Two of this work describes the development of a transposon-mediated insertion mutagenesis strategy to construct a library of *H. influenzae* Rd KW20 clones, each containing a single, characterized insertion, for use in determining the function of annotated genes. In Chapter Four, the library is used to search for new genes mediating genomic transformation. Additionally, in Chapter Four, *in silico* predictions, Q-PCR and directed mutagenic studies are used to further characterize the development of competence and transformation in this species.

CHAPTER TWO

Transposon mutagenesis of *Haemophilus influenzae* Rd KW20

ABSTRACT

The publication of the *Haemophilus influenzae* Rd KW20 genomic sequence in 1995 led to a unique opportunity to study this organism (27). A transposon-based strategy was designed to facilitate the generation of insertions in each of the predicted protein coding sequences in the Rd KW20 genome. The strategy included the creation of specialized cloning vectors to allow post-cloning removal of non-essential plasmid sequences in order to increase the efficiency of recovery of chromosomal insertions. Chromosomal fragments from Rd KW20 were cloned into these specialized vectors using blue-white screening, and the *lacZα* sequences were then removed prior to, or following transposon mutagenesis. The availability of the Rd KW20 genomic sequence facilitated rapid mapping of the transposon insertion sites. Analysis of insertion site characteristics identified a previously unrecognized sequence content bias for a commercially available *in vitro* Tn5 mutagenesis system.

INTRODUCTION

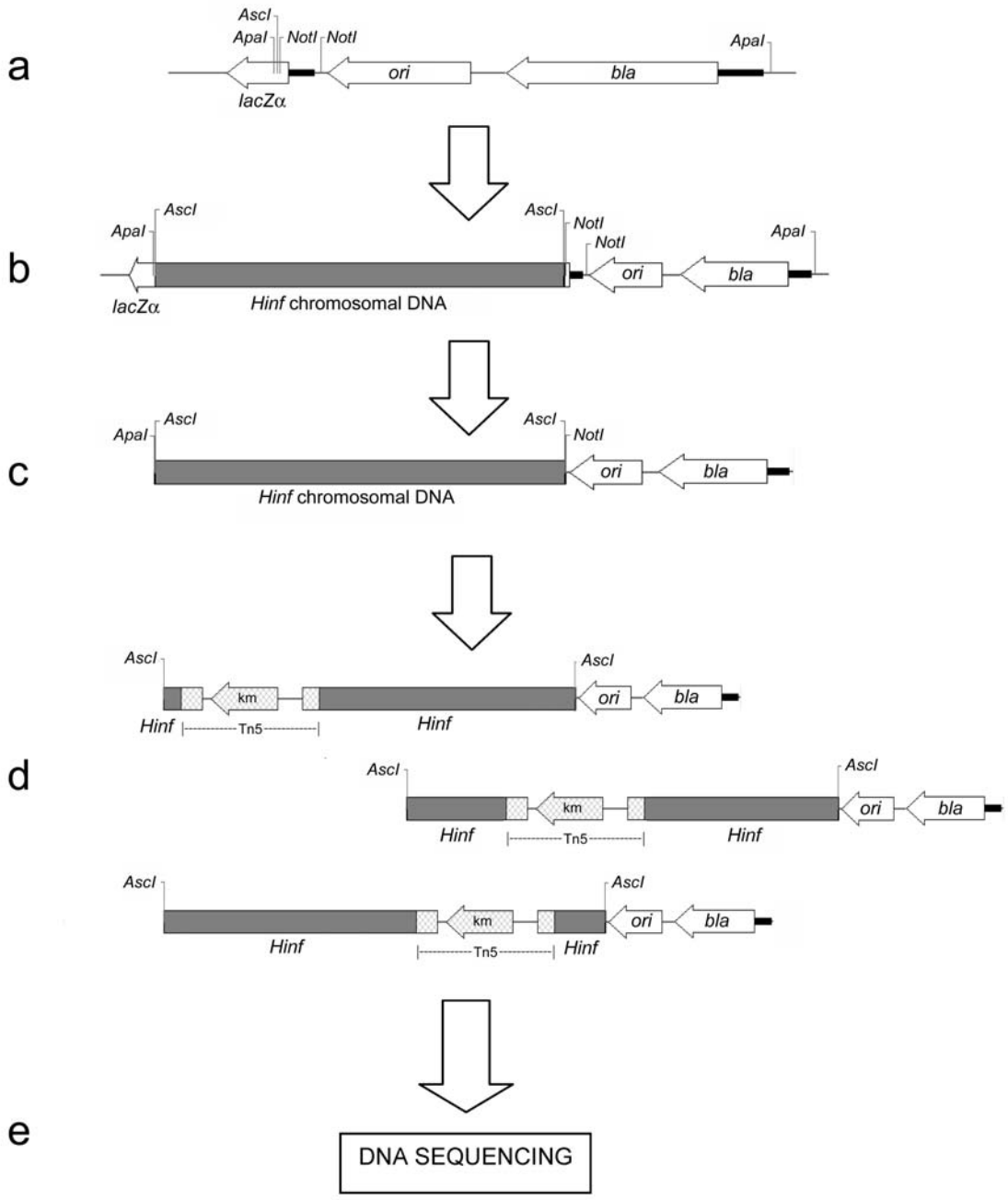
Haemophilus influenzae is a small, pleiomorphic, Gram-negative, facultative anaerobe that can be classified based on the presence or absence of capsular polysaccharide, serotype a-f or non-typeable (NTHi) respectively. Capsulated strains of *H. influenzae*, in particular type b (Hib), can cause serious invasive diseases that include meningitis, septicemia, epiglottitis, cellulitis, septic arthritis, pneumonia and empyema

(10). A vaccine based on the Hib polysaccharide capsule has radically reduced the incidence of disease caused by Hib (1). NTHi is the etiological agent of several human diseases including otitis media and respiratory infections in patients with predisposing conditions, including cystic fibrosis and chronic obstructive pulmonary disease (COPD) (28, 87, 102).

At the commencement of this work in 1997, many techniques that are now commonplace in functional genomics were in their infancy. We chose to utilize a transposon-based mutagenesis strategy to characterize the 1.83-Mbp *H. influenzae* Rd KW20 genome. While transposon mutagenesis was a common procedure at that time, recently introduced *in vitro* transposition systems made this technique more practical. In addition, the availability of the Rd KW20 genomic sequence allowed easier and more rapid mapping of the insertion sites than previously attainable. Strategies involving transcriptomics and proteomics can be useful in defining which products are preferentially expressed in differing conditions, but determination of the role or function of gene products often requires studying the changes in phenotype that result from the elimination of the gene. A directed, PCR-based approach to systemically delete each of the 1700+ annotated protein coding sequences (CDSs) would be capable of providing less ambiguous results than insertion-derived mutants but would have been prohibitively costly given the resources that were available. Thus, a transposon mutagenesis strategy was determined to be the most reasonable approach. The overall strategy for the systematic creation of a *H. influenzae* transposon insertion library is shown in Figure 2.1. In short, fragments of Rd KW20 genomic DNA were cloned into pUC-based plasmid vectors, randomly mutagenized by *in vitro* transposition, and transformed into *E. coli*

DH5α. Plasmids containing transposon insertions were purified from the host bacteria and the sites of insertion were determined by direct DNA sequencing. Preliminary results of the work indicated a large percentage of insertions mapped within the vector sequence. This necessitated the construction of a series of minimal plasmids to attempt to alleviate the observed bias for recovery of vector insertions (see results). In a later phase of this work, insertions that mapped within chromosomal DNA were introduced into *H. influenzae* via its natural transformation mechanism and mutants were screened for defects in transformation ability.

Figure 2.1 – Construction of the *Haemophilus influenzae* Rd KW20 insertion libraries. a-b. Restriction digestion or sonicated fragments of the *Haemophilus* genome were cloned into the *Asc*I site of the pASC series of cloning vectors. **c.** Non-essential regions of the vector backbone were removed by sequential digestion/religation steps with *Apa*I and *Not*I. **d.** The library was mutated by *in vitro* mutagenesis using the transposons Tn7 or Tn5. **e.** The site of transposon insertion was mapped by dideoxynucleotide sequencing from unique primers within the transposon across the insertion junction. In later libraries, steps C and D were reversed (post-mutagenesis minimalization) or chromosomal DNA was cloned directly into a minimal vector lacking the *lacZα* sequence (pASC18MIN). Abbrieviations used: *lacZα,* β-galactosidase gene; *ori,* ColE1 origin of replication; *bla,* ampicillin resistance gene; *km,* kanamycin resistance gene


a
AscI
ApaI
NotI
NotI
ApaI
ori
bla
lacZα
b
ApaI
AscI
AscI
NotI
NotI
ApaI
ori
bla
lacZα
Hinf chromosomal DNA
c
ApaI
AscI
AscI
NotI
ori
bla
Hinf chromosomal DNA
d
AscI
AscI
km
ori
bla
Hinf
Tn5
Hinf
AscI
AscI
km
ori
bla
Hinf
Tn5
Hinf
AscI
AscI
km
ori
bla
Hinf
Tn5
Hinf
e
DNA SEQUENCING

MATERIALS AND METHODS

Strains and plasmids. All *H. influenzae* and *E. coli* strains, and cloning vectors used in this study are listed in Table 2.1.

Enzymes and chemicals. All enzymes were purchased from New England Biolabs (NEB), Beverly, Mass. or Promega, Madison, Wisc., and reactions were carried out as recommended by the manufacturer. *In vitro* transposition kits were purchased from Epicentre Technologies, Madison, Wisc. (Tn5-based) or NEB (Tn7-based). Reagents for DNA sequencing reactions were purchased from United States Biochemical (USB) Corporation, Cleveland, Ohio. All bacterial growth media was obtained from BBL/Difco (Becton Dickenson, Sparks, Md.). All other chemicals were obtained from Sigma-Aldrich, St. Louis, Mo. unless otherwise noted.

Oligonucleotides and linkers. Oligonucleotide linkers were purchased from NEB. Custom oligonucleotides were purchased from Life Technologies, Gaithersburg, Md. Custom oligonucleotides and linkers used to construct the minimal vectors and to modify chromosomal DNA for vector insertion are listed in Table 2.2. Oligonucleotides were resuspended in TE buffer (10 mM Tris-Cl, pH 8.0, 1 mM EDTA) at a concentration of 50 μM.

Ligation reactions. All ligation reactions were performed as recommended by the manufacturer. Ligation of restriction fragments into plasmids was performed by addition of an approximate 4:1 molar ratio of insert to vector. Ligation reactions were performed overnight at 15°C unless otherwise noted.

Table 2.1– Strains and plasmids used in this work

Strain or plasmid	Relevant characteristics	Source or ref.
Strains		
E. coli [a]		
DH5α	F$^-$ Φ80*lacZ*Δ*M15* Δ(*lacIZYA-argF*)*U169 recA1 deoR hsdR17*(r_k^-, m_k^+) *phoA supE44 thi-1 gyrA96 relA1 endA1*	Invitrogen
H. influenzae		
Rd KW20	Capsule-deficient type d derivative, sequenced strain (27)	ATCC (6)
Plasmids		
pUC19	2.7-kbp, P*lac lacZα* Ap^r ColE1 ori	USB
pUC19H°	pUC19 Δ*Hin*dIII	This study
pASC1	pUC19H°, replacement of *Pst*I with *Asc*I	This study
pASC10	2.4-kbp, pASC1 with a 274-bp deletion upstream of *lacZα* and replacement with a *Not*I site	This study
pASC13	2.4-kbp, pASC10 with *Apa*I and *Not*I sites for post-cloning removal of *lacZα* sequences	This study
pASC0299	pASC13::Tn7Cm at nucleotide 1300	This study
pASC15	2.3-kbp, pASC300 ΔTn7, replacement of 107-bp between *bla* and *ori* with *Pst*I site	This study
pASC15MIN	1.7-kbp, pASC15 Δ*Apa*I Δ*Not*I fragments	This study
pASC18	2.3-kbp, pASC15 with tR2 terminator from phage λ inserted upstream of *Asc*I	This study
pASC18MIN	1.7-kbp, pASC18 Δ*Apa*I Δ*Not*I fragments	This study
pASC23	1.7-kbp, pASC15MIN with *Bgl*I and *Bse*RI sites inserted at *Pst*I	This study
pASC30	1.6-kbp, pASC23 with deletion of 34-bp between *bla* and *ori*	This study

[a] Antibiotic concentrations used for selection of *E. coli*: ampicillin (Ap) 100 μg/ml, kanamycin (Km) 50 μg/ml, chloramphenicol (Cm) 50 μg/ml

Table 2.2 – Nucleotide sequences of primers and adapters used in this work

Primer name	Sequence (5' to 3')
KAN-2-FP1	ACCTACAACAAAGCTCTCATCAAC
KAN-2-RP1	GCAATGTAACATCAGAGATTTTGAG
Primer N	ACTTTATTGTCATAGTTTAGATCTATTTTG
Primer S	ATAATCCTTAAAAACTCCATTTCCACCCCT
AP-C1	CTGCACCTCCTCGTTACCAATGCTTAATCAGTGAG
AP-C2	GTGCCTCACTGATTAAGCATTGGTAACGAGGAGGTGCAGTGCA
TERMA	CGCGCCACTGGCGTGCCTTTTTTTGC
TERMB	GGCCGCAAAAAAGGCACGCCAGTGG
LINK1	CGCGGCTCTTCGTGCAGTCACGTCTCG
LINK2	CTAGGCTCTGCACTGACGTGCTTCTCG

Preparative agarose gel electrophoresis. DNA fragments were purified by isolation using agarose gel electrophoresis. The desired bands were removed from the gel and purified by heating at 95°C for 10 minutes to liquefy the agarose followed by two serial extractions with an equal volume of Tris-buffered phenol (pH 7.5). Residual phenol was removed from the aqueous layer by extraction with an equal volume of 2-butanol. The DNA was precipitated by addition of NaCl to a final concentration of 0.5 M and two volumes of 95% ethanol. The pellet was washed with 70% ethanol, dried and resuspended in TE buffer.

Transformation of plasmids into *E. coli*. Plasmid constructs were transformed into *E. coli* DH5α by electroporation. Electrocompetent cells were prepared by the method of Sharma and Schimke (88). A portion of ligation reactions (0.5 - 1.0 µl) was added to 75 µl of ice-cold competent cells and electroporated at 17,000 v/cm using an Eppendorf Electroporator 2510 (Brinkmann Instruments, Westbury, NY). Electroporated cells were immediately transferred to 1 ml of SOC broth (0.5% Yeast Extract, 2% Tryptone, 10 mM NaCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose) and plated on LB agar (1% Tryptone, 0.5% Yeast Extract, 170 mM NaCl, 1.5% Bacto-agar) containing appropriate antibiotics for selection and 40 µg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) when appropriate for screening for insertions into the vector. Antibiotic concentrations used for selection of *E. coli* transformants are listed in Table 2.1.

Purification of plasmids. A modification of the alkaline lysis technique was utilized to isolate plasmids from *E. coli* DH5α (62). Colonies displaying the correct phenotype were subcultured into 5 ml Terrific broth (TB) (1.2% Tryptone, 2.4% Yeast

Extract, 0.4% v/v Glycerol, 72 mM K_2HPO_4, 17 mM KH_2PO_4) (98), containing the appropriate antibiotic concentrations and incubated with vigorous shaking overnight at 37°C. Cells were collected by centrifugation at 14,000 x g for 5 minutes. The supernatant was aspirated and the cells resuspended in 100 µl Solution 1 (50 mM Glucose, 25 mM Tris-HCl, pH 8.0, 10 mM EDTA). Once resuspended, 200 µl of Solution 2 (0.2 N NaOH, 1% SDS) was added and the solution was mixed by gentle inversion of the tube and incubated on ice for 5 minutes. The mixture was neutralized by addition of 150 µl of Solution 3 (3 M K^+, 5 M acetate), mixed by vortexing, and incubated on ice for 5 minutes. Cellular debris was removed by centrifugation at 14,000 x g for 5 minutes and the supernatant extracted with an equal volume of Tris-buffered phenol (pH 7.5). Residual phenol was removed from the aqueous phase by two extractions with equal volumes of 2-butanol. The plasmid DNA was precipitated by addition of one-tenth volume of 5 M NaCl and two volumes of 95% ethanol and collected by centrifugation at 14,000 x g for 5 minutes. The pellet was washed with 70% ethanol and allowed to dry. The DNA was resuspended in 50 µl TE buffer and RNA digested by incubation with 10 µg of RNase for 10 minutes at room temperature. The phenol and 2-butanol extractions were omitted for plasmids that were to be directly sequenced.

Construction of minimalized cloning vectors. A series of specialized cloning vectors was developed to lower the frequency at which insertions were recovered and mapped in vector sequences. The parent plasmid used in the construction was pUC19. The major steps undertaken to create this series of plasmids are described below and represented in Figure 2.2.

Figure 2.2 – Development of the pASC family of minimal cloning vectors. The major intermediates in the development of the pASC vectors are shown. The white arrows indicate the β-galactosidase coding region and the black arrows indicate the β-lactamase coding region (*bla*). The gray arrows indicate the approximate position of the plasmid origin of replication. -35 *bla* indicates the position of the start of the *bla* promoter region. Nonessential sequence is defined as all vector sequences outside of the origin of replication, *bla* and the *bla* promoter region. The Tn7 site notated in pASC13 indicates the site of the Tn7 insertion in pASC0299 used for deletion of nonessential sequence between the *bla* terminus and the origin region.

[a] Indicates insertion frequency with Tn7.

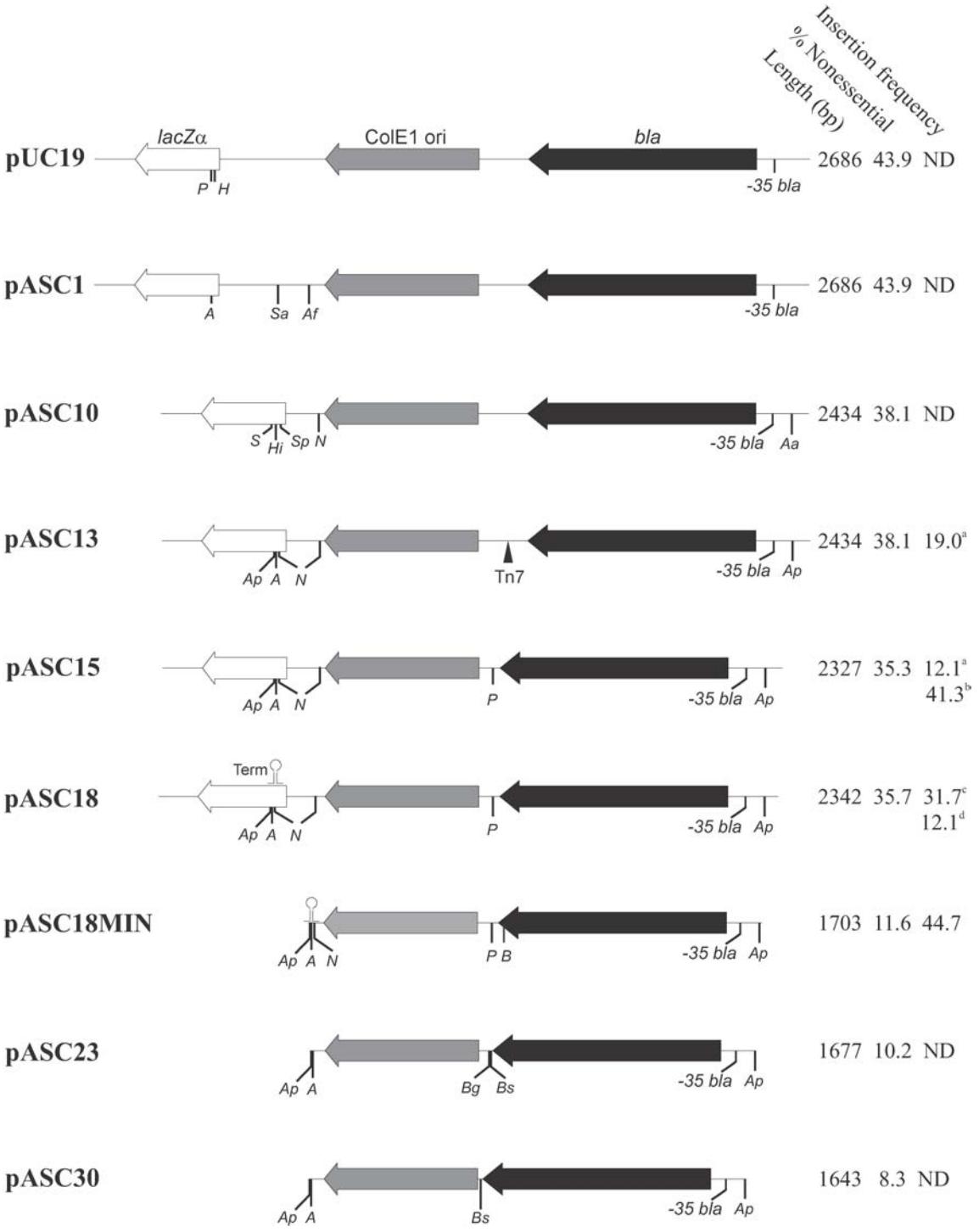
[b] Indicates insertion frequency with Tn5.

[c] Indicates insertion frequency with pre-mutagenesis minimalization of the vector.

[d] Indicates insertion frequency with post-mutagenesis minimalization of the vector.

ND indicates the vector not utilized for library creation.

Restriction enzyme abbreviations used: A=*Asc*I; Aa=*Aat*II; Af=*Afl*III; Ap=*Apa*I; B=*Ban*I; Bg= *Bsg*I; Bs=*Bse*RI; H=*Hin*dIII; Hi=*Hinc*II; N=*Not*I; P=*Pst*I; S=*Sma*I; Sa=*Sap*I; Sp=*Sph*I


Length (bp)
% Nonessential
Insertion frequency
pUC19
lacZα
ColE1 ori
bla
P H
-35 bla
2686 43.9 ND
pASC1
A
Sa Af
-35 bla
2686 43.9 ND
pASC10
S Hi Sp N
-35 bla Aa
2434 38.1 ND
pASC13
Ap A N
Tn7
-35 bla Ap
2434 38.1 19.0[a]
pASC15
Ap A N
P
-35 bla Ap
2327 35.3 12.1[a] 41.3[b]
pASC18
Term
Ap A N
P
-35 bla Ap
2342 35.7 31.7[c] 12.1[d]
pASC18MIN
Ap A N
P B
-35 bla Ap
1703 11.6 44.7
pASC23
Ap A
Bg Bs
-35 bla Ap
1677 10.2 ND
pASC30
Ap A
Bs
-35 bla Ap
1643 8.3 ND

a. pUC19H° - The *Hin*dIII site within the multiple cloning region (MCR) of pUC19 was deleted by restriction digestion with *Hin*dIII followed by removal of the single-stranded overhangs by Mung Bean nuclease digestion. The vector was recircularized by blunt-end ligation using T4 DNA ligase to create pUC19H°.

b. pASC1 – pUC19H° was digested with *Pst*I and treated with Mung Bean nuclease. Phosphorylated *Asc*I linkers (NEB) were ligated to the blunt-ended plasmid and excess linkers were removed by preparatory agarose gel electrophoresis. Following overdigestion with *Asc*I, the vector was circularized by treatment with T4 DNA ligase, resulting in pASC1.

c. pASC10 - Deletion of non-essential vector sequence between the replication origin priming site and the *lac* promoter was performed by digestion of pASC1 at the unique *Sap*I site and treatment with BAL-31 nuclease. Phosphorylated *Not*I linkers (NEB) were ligated to the plasmid and excess linkers removed by preparative agarose gel electrophoresis. The plasmid was circularized by overdigestion with *Not*I followed by treatment with T4 DNA ligase. This process was repeated at the unique *Afl*III site in the plasmid. The intervening sequence between the two deletion sites was removed during the recirculation of the large vector fragment forming pASC10.

d. pASC13 - A cloning vector was created by introducing paired restriction sites flanking the *Asc*I cloning site in pASC10 to allow both blue-white screening and minimalization of the vector backbone by excision of non-essential sequences following the cloning step. The *Aat*II site present immediately upstream of the *bla* promoter region was replaced by digestion with *Aat*II, treatment with Mung Bean

nuclease and ligation of phosphorylated *Apa*I linkers (NEB). Excess linkers were removed by preparative electrophoresis and the plasmid circularized by overdigestion with *Apa*I and ligated with T4 DNA ligase to form pASC11 (not shown). An *Apa*I site was introduced immediately downstream of the *Asc*I site in pASC11 by digestion with *Hinc*II and *Sma*I followed by ligation of phosphorylated *Apa*I linkers as previously described to create pASC12 (not shown). pASC13 was created by introducing a second *Not*I site in pASC12 upstream of the *Asc*I site. This was performed by digestion with *Sph*I, treatment with Mung Bean nuclease and ligation of *Not*I linkers as described above.

e. pASC15 – pASC13 retained a significant amount of non-essential sequence between the terminus of *bla* and the start of the origin region. A portion of this region was removed in two steps by utilizing a Tn7 insertion recovered in pASC13 (at pUC19 coordinate 1555). The plasmid (pASC0299) was linearized at the unique *Bgl*II site present in the Tn7 element and treated with BAL-31 to delete the neighboring sequence. A *Pst*I linker was ligated to the treated plasmid and the vector was circularized by overdigestion with *Pst*I followed by ligation with T4 DNA ligase. This procedure was repeated at the *Spe*I site at the other Tn7 terminus. The intervening transposon sequence between the *Pst*I sites was removed during this process.

f. pASC18 – A Rho-independent transcription terminator (tR2 from phage λ) was added to pASC15 to prevent transcription from inserted DNA interfering with plasmid maintenance activities. pASC18 was created by linearizing pASC15 by an *Asc*I and partial *Not*I digestion and inserting a double-stranded cassette formed

by the oligonucleotides TermA and TermB (Table 2.2). The vector was circularized by ligation with T4 DNA ligase.

g. pASC18MIN – A minimized version of pASC18 was also used for direct cloning of chromosomal fragments. In order to remove the non-essential region between the *Asc*I cloning site and the promoter region of *bla,* pASC18 was minimized by digestion with *Apa*I followed by circularization with T4 DNA ligase. Similarly, the region between the *Asc*I cloning site and the replication origin was removed by digestion with *Not*I.

h. pASC23 – Further minimization of the pASC-based cloning vectors was performed in multiple steps. pASC15 was cleaved by *Not*I digestion, treated with Mung Bean nuclease to remove the single-stranded overhangs and the large fragment purified by agarose gel electrophoresis. The vector was circularized by blunt-ended ligation. The plasmid was cleaved at the *Apa*I sites, the large fragment purified by gel electrophoresis and recircularized. In order to maximize the removal of non-essential sequence between the end of *bla* and the uncharacterized start of the *ori* region, a strategy developed by Ariazi and Gould was employed. A double-stranded adaptamer formed by the oligonucleotides AP-C1 and AP-C2 (Table 2.2) was ligated into a *Pst*I and *Ban*I digested vector to form pASC23. This introduced the class IIS restriction sites *Bsg*I and *Bse*RI to facilitate unidirectional deletions of 14-bp of plasmid sequence from a fixed position at the end of the *bla* gene.

i. pASC30 – pASC30 was created by utilizing the introduction of *Bsg*I and *Bse*RI into the pASC-based vector. pASC23 was digested with *Bsg*I followed by heat-

inactivation of the enzyme. The digested plasmid was treated with *Bse*RI, heat-inactivated and treated with Mung Bean nuclease to remove the 3'-overhangs. The treated plasmid was circularized by treatment with T4 DNA ligase and transformed into electrocompetent DH5α. This process was repeated until no further transformants were recovered following a round of deletion. The final plasmid recovered was designated pASC30 and the extent of deletion in this region was determined by automated DNA sequencing.

Construction of chromosomal libraries. All libraries were constructed by cloning *H. influenzae* chromosomal DNA into the *Asc*I site of the various pASC derivatives. Vectors were digested with *Asc*I, as recommended by the manufacturer, and the reactions terminated by addition of one volume of Tris-buffered phenol (pH 7.5). The reaction was centrifuged and the aqueous layer was removed and extracted twice with 2-butanol followed by ethanol precipitation. Terminal phosphates were removed from the linearized vector by treatment with Calf Intestinal Alkaline Phosphatase (CIP) as directed by the manufacturer. The CIP-treated vector was subjected to ligation overnight at 15°C to circularize any partially phophorylated molecules. The linear, monomer form of the plasmid was purified by agarose gel electrophoresis and used in the construction of the libraries.

In order to obtain purified chromosomal DNA, *Haemophilus influenzae* Rd strain KW20 was grown to saturation in Brain Heart Infusion supplemented with 10mg/L hemin and 1mg/L β-NAD (sBHI) and collected by centrifugation at 4000 x g for 10 minutes. The chromosomal DNA was isolated and purified by separation on a cesium

chloride gradient using the large-scale preparation method described by Wilson (11). The DNA was fragmented by sonication or restriction digestion and blunted-ended, if necessary, by treatment with Mung Bean nuclease. Restriction digest libraries were created using *Fsp*I, *Pvu*II, *Eco*RI, *Xmn*I, *Swa*I, or *Psi*I. Phosphorylated *Asc*I linkers were ligated to the target DNA at 15°C overnight. Excess linkers were removed by preparative agarose gel electrophoresis and chromosomal fragments greater than 2-kbp were recovered by agarose gel electrophoresis. Four micrograms of linker-treated chromosomal DNA were incubated with *Asc*I to generate overhangs and incubated with 1 μg of dephosphorylated vector and T4 DNA ligase in a 50 μl reaction. A portion of the ligation mixture (0.5 – 1 μl) was transformed into DH5α by electroporation and recombinant plasmids were recovered. Residual non-recombinant vectors were removed from the libraries by preparative gel electrophoresis.

The plasmid pASC18MIN, which lacks the *lacZα* gene, was also used to make a mutagenesis library. *H. influenzae* Rd KW20 chromosomal DNA was digested with *Bgl*II and ligated to the double-stranded adapter LINK1•2 (Table 2.2). This adapter has an overhang to allow ligation to the end of the target DNA but removes the *Bgl*II recognition site. Ligation was performed at high target concentration in *Bgl*II buffer (50 mM Tris-Cl, pH 7.9, 100 mM NaCl, 10 mM $MgCl_2$, final concentration) supplemented with 1mM ATP, 3 mM DTT, 10 units of *Bgl*II, and 2.0 units of T4 DNA ligase. Ligation temperatures were cycled 150 times (15°C for 15 min, 30°C for 1 min) as described by Lund *et al.* (61). This method favors the addition of linkers over fusions between chromosomal fragments as the latter retain a *Bgl*II site following the low-temperature ligation cycle and are recleaved during the high-temperature incubation. Excess linkers

were removed by preparative gel electrophoresis. The linker-treated DNA was cloned into pASC18MIN using the same methodology. The other terminus of the LINK1•2 adapter had an *Asc*I-compatible overhang. Ligation of the target DNA into *Asc*I digested vector would remove the *Asc*I site and prevent recleavage while a recircularized vector reconstitutes a full *Asc*I site and would be cleaved during the high-temperature cycles. Ligations were performed in *Asc*I buffer (50 mM potassium acetate, 20 mM Tris acetate, pH 7.9, 10 mM magnesium acetate, final concentration) containing 1 mM ATP, 3 mM DTT, 10 units of *Asc*I, and 2 units of T4 DNA ligase and cycled as above. Following ligation, 0.5µl of ligation mixture was used to transform DH5α by electroporation and recombinant plasmids recovered as previously described.

Post-cloning minimalization by lacZ removal. The regions of the vector containing the *lacZα* coding region and promoter, while useful to screen for recombinants, provide a large region that is a target for transposon insertions. Prior to mutagenesis, these expendable regions were removed from the vector by restriction digestion. The region between the *Asc*I cloning site and the *bla* promoter was removed by digestion with *Apa*I and preparative gel electrophoresis employed to resolve the recombinants from the small *Apa*I fragment. The fraction of the gel containing recombinant plasmid fragments was removed from the gel, and DNA was purified as previously described. The plasmids were recircularized by ligation at low concentration (< 2 µg/ml) and transformed into DH5α by electroporation. The recombinant plasmids were harvested from ampicillin-resistant colonies and subjected to deletion of the *Not*I fragment, comprising the region between the *ori* terminus and *Asc*I site, using the same protocol. Since the recognition sites for *Apa*I and *Not*I are rare in the Rd KW20 genomic

sequence, these manipulations would not have lead to the loss of a significant number of plasmid clones.

***In vitro* transposon mutagenesis.** The recombinant *H. influenzae* Rd KW20 libraries were mutagenized using two *in vitro* transposition kits. Initially, the Genome Priming System (GPS-1) from NEB was employed following the manufacturer's instructions. The kit is composed of a Tn7-based transposase (TnsABC complex) with a gain-of-function mutation that reduces insertion bias (95) and a donor plasmid that carries a mini-Tn7 transposon with a chloramphenicol or kanamycin resistance marker, pGPS1.1 and pGPS2.1 respectively. An 18 µl reaction contained GPS buffer (25 mM Tris-Cl, pH 8.0, 2 mM ATP, 2 mM DTT, final concentration), 0.02 µg donor plasmid, and approximately 0.1 µg of target DNA. One µl of the TnsABC transposase complex was added to the mixture and incubated for 10 min at 37°C to allow assembly of transposase-transposon complexes. The transposition reaction was initiated by the addition of 15 mM MgOAc and the reaction incubated for 1 hour at 37°C. The reaction was terminated by incubation at 75°C for 10 min and a portion of the reaction (0.5 – 1.0 µl) was used to transform DH5α by electroporation. Chloramphenicol-resistant colonies were subcultured into TB-Cm and, following overnight growth, plasmids were recovered as previously described.

The second *in vitro* mutagenesis system employed was the EZ::Tn <KAN-2> kit from Epicentre Technologies. This is a Tn5-based system that uses a hyperactive mutant of the Tn5 transposase (39) and a mini-Tn5 transposon that carries a kanamycin resistance marker. A 10 µl reaction contained EZ::TN reaction buffer (50 mM Tris-OAc, pH 7.5, 150 mM KOAc, 10 mM MgOAc, 4 mM spermidine, final concentration), 0.2 µg

target DNA, 0.1 pmol EZ::TN <KAN-2> transposon, and 1 unit EZ::TN transposase. The reactions were incubated for 2 hours at 37°C to allow transposition and terminated by addition of SDS to 0.1% with incubation at 70°C for 10 min. A 0.5 – 1.0 µl sample of the reaction was electroporated into DH5α and mutagenized plasmids were recovered as previously described.

Insertion mapping. The locations of transposon insertion sites were mapped by sequencing out from the transposon unit into the flanking DNA. This was accomplished using primers that annealed to unique binding sites located near the transposon termini. Sequencing reactions were performed using the Thermo Sequenase Radiolabeled Terminator Cycle Sequencing kit from USB Corporation. For each mapped plasmid, four sequencing reactions were performed with each containing a different ^{33}P-ddNTP for sequence termination. Each 7.5 µl reaction contained 1 µl (~100ng) plasmid DNA, 12.5 pmol sequencing primer, 15 pmol dATP, dGTP, dCTP, dTTP, 0.5 µl reaction buffer (260 mM Tris•HCl, pH 9.5, 65 mM $MgCl_2$), 0.5 µl polymerase mix (2 units Thermosequenase, 0.0003 units *Thermoplasma acidophilum* inorganic pyrophosphatase, 50 mM Tris•HCl, pH 8.0, 1 mM DTT, 0.1 mM EDTA, 0.5% Tween 20, 0.5% Nonidet P-40, 50% glycerol), and 0.3 pmol [α-^{33}P] ddNTP (1500Ci/mmol). The reactions were cycled 30 times (95°C, 30s; 55°C, 30s; 72°C, 60s) and terminated by addition of 4 µl Stop Solution (95% formamide, 20 mM EDTA, 0.05% bromphenol blue, 0.05% xylene cyanol FF). The nucleotide sequences of the oligonucleotide primers used for mapping Tn7 insertions (Primer N and Primer S) and Tn5 insertions (KAN-2-FP1 and KAN-2-RP1) are listed in Table 2.2. The reactions were resolved by TBE-buffered PAGE on a Sequi-Gen GT Sequencing cell (Bio-rad Laboratories, Hercules, Calif.). The 0.4 mm sequencing gels

were freshly cast and composed of 8M urea (Certified ACS, Fisher Chemicals, Fairlawn, NJ), 8% Acrylamide:Bis-acrylamide 29:1 (Fisher Chemicals) and TBE (89 mM Tris-borate, pH 8.3, 2.0 mM EDTA) and polymerized by addition of 0.1% APS (ammonium persulfate, Fisher Chemicals) and 0.025% TEMED (N,N,N',N'-tetramethylethylenediamine, Bio-rad Laboratories). The sequencing gels were run for 15 minutes at 2000 V prior to the loading of samples. The sequencing reactions were incubated at 70°C for 5 minutes prior to loading 5 µl into the appropriate sequencing well. The gels were run at a constant 2000 V until the bromphenol blue dye reached the bottom of the gel. The urea was removed from the gel by soaking for 30 minutes in a fixing solution composed of 5% v/v acetic acid and 20% v/v methanol. The gels were transferred onto Whatmann 3MM paper and dried in a gel dryer (Bio-rad Laboratories, model 583), followed by exposure to autoradiography film overnight in order to visualize the bands. Query sequences of 20+ bp, derived from the sequence reads, were analyzed using a custom, python-based program (Mapper, unpublished data) that allowed for rapid comparison to the *H. influenzae* genomic sequence and vector sequences to determine the precise location of transposon inserts (Figure 2.3). Mapper also allowed deconvolution of multiple sequences in the same sequencing reaction using the genomic and vector DNA sequences as a filter (Figure 2.4). This feature is an invaluable resource in situations where multiple insertions are recovered in the same plasmid. We were able to deconvolute and map the insertions in most plasmids containing two transposon insertions and many of those containing three transposon insertions.

Figure 2.3 – Single Primer Mapping of Insertion Sites. A. A generalized representation of the Tn7 transposon insertion into a *H. influenzae* Rd KW20 chromosomal fragment. Either sequencing primer, Primer N or Primer S, can be used to map the site of the insertion. Primer N anneals within the Tn7R terminal end and sequences outward into the chromosomal fragment. Primer S anneals within the Tn7L terminus. **B.** Autoradiograph from the sequencing of pASC0282. Lanes are terminated from left to right with dideoxy- guanine, adenine, thiamine and cytosine. **C.** The location of the sequence is obtained by entering the query sequence into the python-based Mapper program. Mapper compares the input sequence against the positive or negative strands of any chosen genome sequence. For pASC0282, a query sequence of 10 nucleotides was sufficient to locate the site of insertion at nucleotide position 984122 of the positive strand of the Rd KW20 genomic sequence. This nucleotide position is located within the CDS HI0925, which encodes for a conserved hypothetical protein.

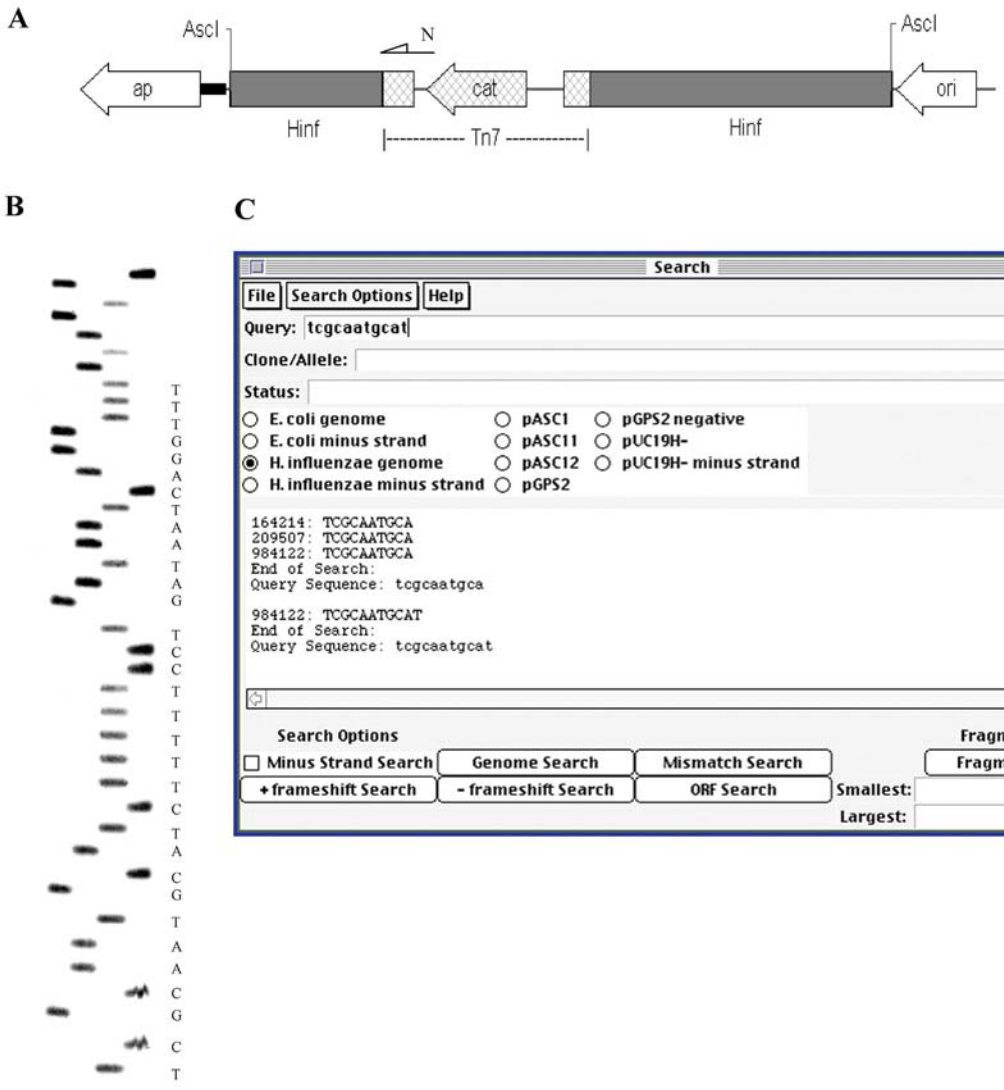
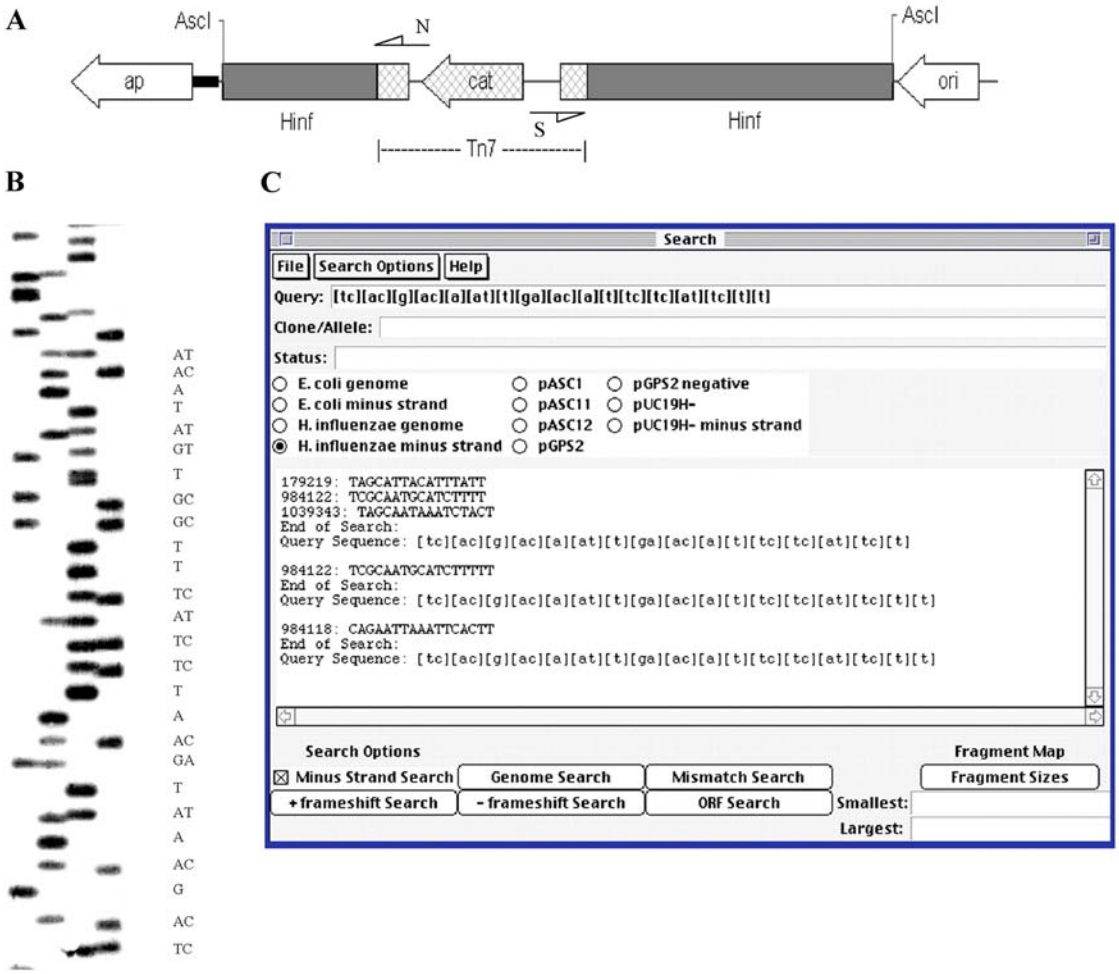

A
AscI
N
AscI
ap
cat
ori
Hinf
Tn7
Hinf
B
C
Search
File
Search Options
Help
Query: tcgcaatgcat
Clone/Allele:
Status:
E. coli genome
E. coli minus strand
H. influenzae genome
H. influenzae minus strand
pASC1
pASC11
pASC12
pGPS2
pGPS2 negative
pUC19H-
pUC19H- minus strand
164214: TCGCAATGCA
209507: TCGCAATGCA
984122: TCGCAATGCA
End of Search:
Query Sequence: tcgcaatgca
984122: TCGCAATGCAT
End of Search:
Query Sequence: tcgcaatgcat
Search Options
Minus Strand Search
Genome Search
Mismatch Search
+ frameshift Search
- frameshift Search
ORF Search
Fragment Map
Fragment Sizes
Smallest:
Largest:

Figure 2.4 – Dual primer convoluted sequence mapping. A. A generalized representation of the Tn7 transposon insertion into a *H. influenzae* Rd KW20 chromosomal fragment. Both sequencing primers, Primer N and Primer S, are used simultaneously to map the site of the insertion. Alternatively, a single primer is utilized to map the sites of two transposon insertions in the same plasmid. **B.** Autoradiograph from the sequencing of pASC0282. Lanes are terminated from left to right with dideoxy- guanine, adenine, thiamine and cytosine. One can see the convolutions in the sequence data where two nucleotides migrate to the same position in the gel. **C.** The locations of the sequences are obtained by entering the query sequence into the Mapper program. The query sequence is derived by bracketing the two sequence possibilities at each convoluted site. For convoluted sequences, Mapper computes all the potential combinations and compares them against the positive or negative strand of any chosen genome sequence. For pASC0282, a query sequence of 17 nucleotides was sufficient to locate the site of insertion. The sequence from Primer N indicates an insertion at nucleotide position 984122 of the positive strand of the Rd KW20 genome. The sequence from Primer S indicates an insertion at position 984118. The discrepancy is due to the duplication of the target site during transposition. These nucleotide positions were in the CDS HI0925.


A
AscI
N
AscI
ap
cat
ori
Hinf
S
Hinf
|---------- Tn7 ----------|
B
AT
AC
A
T
AT
GT
T
GC
GC
T
T
TC
AT
TC
TC
T
A
AC
GA
T
AT
A
AC
G
AC
TC
C
Search
File
Search Options
Help
Query: [tc][ac][g][ac][a][at][t][ga][ac][a][t][tc][tc][at][tc][t][t]
Clone/Allele:
Status:
E. coli genome
E. coli minus strand
H. influenzae genome
H. influenzae minus strand
pASC1
pASC11
pASC12
pGPS2
pGPS2 negative
pUC19H-
pUC19H- minus strand
179219: TAGCATTACATTTATT
984122: TCGCAATGCATCTTTT
1039343: TAGCAATAAATCTACT
End of Search:
Query Sequence: [tc][ac][g][ac][a][at][t][ga][ac][a][t][tc][tc][at][tc][t]
984122: TCGCAATGCATCTTTTT
End of Search:
Query Sequence: [tc][ac][g][ac][a][at][t][ga][ac][a][t][tc][tc][at][tc][t][t]
984118: CAGAATTAAATTCACTT
End of Search:
Query Sequence: [tc][ac][g][ac][a][at][t][ga][ac][a][t][tc][tc][at][tc][t][t]
Search Options
Minus Strand Search
Genome Search
Mismatch Search
+frameshift Search
-frameshift Search
ORF Search
Fragment Map
Fragment Sizes
Smallest:
Largest:

Analysis of Tn5 insertion site characteristics. The number of Tn5 insertion sites recovered allowed for a detailed examination of insertion site biases. Tn5 insertion site characteristics were analyzed by examining the sequence composition of the 40 bp flanking the chromosomal and vector insertion sites. Although the Tn5 insertions were only sequenced unidirectionally, two factors allow the prediction of the 20-bp located on the other side of most insertion sites. First, *in vitro* Tn5 transposition is precise and always results in the duplication of 9-bp of the target sequence (38, 39). Secondly, the sequence of the chromosomal DNA and the restriction enzymes used were both known, thus the ends of the chromosomal inserts could be fairly well defined. Although it is not discussed in the results, the sequences up to 100-bp from the insertion sites were able to be read. Even single nucleotide deviation from the published Rd KW20 sequence was extremely rare. With these considerations in mind, it is reasonable to assume that errors in these predictions are sufficiently rare to not have introduced biases in the results.

In most cases, it is impossible to determine if multiple copies of recovered insertions were due to independent transposition events or duplication of a parent plasmid during our isolation protocols. To avoid skewing the results, only a single copy of each insertion site was included in the analysis. The determination of single nucleotide effects on insertion site preference was determined by calculating the frequency of individual nucleotides at each position relative to the insertion site. Structural analysis of the nucleic acid properties immediately around the insertion sites was performed as previously described for the examination of P-element insertion sites in *Drosophila melanogaster* and Sleeping Beauty and Tc1/*mariner* transposable elements in human and artificial constructs (59, 104). Twenty nucleotides flanking each insertion site were examined for

various effects including A-philicity (the propensity of the sequence to adopt the A form of DNA), B-DNA twist (the propensity of the sequence content to cause slight twisting of the helix and thus changes in the major and minor grooves), bendability, protein-induced deformability, and tri-nucleotide GC content using a three base sliding window and averaged over all the sequences. To allow comparison of the patterns observed at Tn5 insertion sites, 715 random 40-bp sequences from Rd KW20 and 238 random 40-mers from pASC18 were generated *in silico* and used to determine the average values of these effects for the chromosomal and vector sequences. Values used for the effects of dinucleotide or trinucleotide sequence content on DNA structures were compiled by Liao *et al.* from studies on DNA crystal structures (59).

Additional analysis of the DNA sequences surrounding the Tn5 insertion sites in *H. influenzae* DNA was performed using the bend.it program (http://hydra.icgeb.trieste.it/~kristian/dna/) that plots GC content and bendability propensity calculated using the DNase I based bendability parameters and the consensus bendability scale (19, 31).

RESULTS

Creation of a minimized cloning vector. Preliminary experiments, using mini-Tn10 transposons, indicated a strong insertion bias into the *Hin*dIII site within the MCR of pUC19 (data not shown). This site was deleted by restriction digestion with *Hin*dIII and removal of the single-stranded overhangs to create pUC19H°. The removal of 4 nucleotides created a frameshift mutation in *lacZα* although apparent slippage during transcription resulted in light-blue colonies and allowed continued use of X-gal-based

screening for insertions. In order to simplify cloning of chromosomal fragments and subsequent digestion from the vector, an *Asc*I site was introduced into pUC19H°. The resulting vector, pASC1, includes the replacement of the *Pst*I site with a restriction site (*Asc*I) that occurs rarely in the Rd KW20 genome (5 sites total) and the return of the *lacZα* gene to the proper translational frame.

To further reduce the frequency of insertions into the vector, additional deletions were performed to remove other non-essential plasmid sequences (Figure 2.2). The original pUC19 cloning vector contains, as essential sequence, a ColE1-based origin of replication derived from pBR322 (pMB1) and the *bla* gene and promoter necessary for ampicillin selection. A portion of the β-galactosidase gene (*lacZα*) and associated promoter are present in order to facilitate simple screening for insertions into the plasmid. This sequence becomes nonessential after cloning and represents a large sink for potential transposon insertions. Thus, 1180 of 2686 base pairs (43.9%) of the vector can be considered nonessential to plasmid maintenance. To minimize the nonessential sequence between the origin priming site and the *lac* promoter, pASC1 was cleaved at the unique *Sap*I site and vector sequence deleted by BAL-31 nuclease treatment. A *Not*I site was introduced to the vector by ligation of *Not*I linkers to the BAL-31 treated plasmid. This process was repeated at the unique *Afl*III site present in the plasmid and the two deletions were fused to form pASC10. This deletion resulted in the net removal of 252 bp of nonessential sequence from the plasmid and reducing the nonessential regions to 38.1% of the total sequence.

The vector was further manipulated to allow post-cloning vector minimalization. This was accomplished by the introduction of restriction sites that would continue to

permit the use of blue-white screening but would enable the removal of the nonessential *lacZα* sequences prior to transposon mutagenesis. An *Apa*I site was introduced into the *Aat*II site present immediately upstream of the *bla* -35 box. A second *Apa*I site was introduced immediately downstream of the *Asc*I site within *lacZα*. A second *Not*I site was introduced immediately upstream of the *Asc*I site. The resulting plasmid, pASC13, allowed post-cloning excision of the *Apa*I and *Not*I fragments to remove an additional 639 bp of nonessential sequence. This manipulation leaves 16.1% of the vector as nonessential sequence and available for transposon insertion. pASC13 retained a significant amount of nonessential sequence between the end of the *bla* gene and the start of the replication origin. An additional 107-bp from this region were removed in two steps by utilizing a Tn7-insertion recovered in pASC13 (at pUC19 coordinate 1555). The resulting vector, pASC15 retained 10.8% nonessential sequence following the post-cloning manipulations. Post-cloning minimalization of the plasmids by *Not*I and *Apa*I would have little effect on the population of the clones since the recognition sites for these two enzymes are relatively rare in the Rd KW20 genome (1 and 25 sites, respectively).

Two additional changes were made to pASC15. In the first, pASC18 was created by linearizing the vector by an *Asc*I and partial *Not*I digestion and inserting an intrinsic, Rho-independent terminator (tR2 from phage λ). This was performed to alleviate concerns that the lack of apparent complexity in the libraries used might be due to transcription read-through from promoters within the chromosomal inserts into the plasmid replication origin, thus making those plasmids less fit. The final minimal vectors created in this study were formed by series of modifications to pASC15. First, pASC15

was minimalized by removal of the *Not*I and *Apa*I fragments resulting in a vector (pASC22) without the *lacZα* region and only 100 bp between the *ori* terminus and the *bla* regulatory region. Since the small region of nonessential sequence between the C-terminus of *bla* and the *ori* start continued to be a sink for transposon insertions and the precise location of the replication origin was uncharacterized, a strategy developed by Ariazi and Gould was employed to allow deletion of the region (9). The class IIS restriction sites *Bsg*I and *Bse*RI were introduced immediately downstream of the *bla* stop codon in pASC22 to form pASC23. These sites facilitated unidirectional deletions of 14 bp of plasmid sequence from a fixed position at the end of the *bla* gene. Repeated cycles of deletion were continued into the *ori* region until further deletion events resulted in the loss of plasmid viability. This process was performed on pASC23 and resulted in the removal of 34 additional bp from this region. However, the resulting vector, pASC30, was not utilized in this work as the focus shifted to characterizing the transposon mutants that had already been generated.

Tn7 *in vitro* mutagenesis. Initial results from the use of the GPS-1 mutagenesis system were promising. The first group of plasmids sequenced came from a library derived by cloning sonicated Rd KW20 chromosomal DNA into pASC13 and mutated with Tn7Cm. A total of 52 plasmids were sequenced with the resulting mapping of 58 insertion sites. Of these, 47 were insertions into chromosomal DNA (44 unique insertions) representing 30 CDSs and 7 intergenic sites. There were 11 mapped insertions into pASC13 (5 unique) comprising 19.0% of the total insertions. Seven of these vector insertions were in a single site between the end of the β-lactamase coding region and the start of the replication origin. This site was used in the BAL-31 mediated

deletion of pASC13 that created pASC15. A problem that arose during the initial round of mutagenesis was a tendency of the Tn7 construct to form a cointegrate insertion. The original pGPS2 plasmid provided with the NEB kit included a population of plasmid dimers. These led to the occasional insertion of the entire pGPS2 plasmid and resulted in convoluted sequencing results in which one sequence was derived from the donor plasmid itself. Although we were able to successfully map the locations of the chromosomal or vector insertions associated with these insertions, the frequency of these events, 12 cointegrates in 52 plasmids (23.1%), made the use of this kit less than ideal. Subsequently, NEB released a modified version of the plasmid lacking the concatemers. Using the modified version of the GPS-1 mutagenesis system, Rd KW20 chromosomal DNA was cloned into pASC15 and mutated again with Tn7. Libraries were created using sheared chromosomal DNA and DNA digested with *Fsp*I or *Pvu*II. A total of 93 plasmids were sequenced resulting in 99 mapped insertions. Of these, 87 were located in *H. influenzae* chromosomal DNA. However, these 87 insertions were found at only 29 different sites and in only 5 different CDSs and 3 intergenic regions. There were also 12 insertions into the vector (6 unique) representing 12.1% of mapped insertions. Analysis of the insertions recovered indicated that there appeared to be an insertion hotspot located at chromosomal location 995057, within HI0936, a gene implicated in cytochrome C-type biogenesis. This site represented 33 of the 99 mapped insertions (33.0%). This rate might suggest that the mutagenesis and recovery procedures were inefficient and resulted in the repeated recovery of the same insertion. The two lines of evidence on this are contradictory; this same insertion was prominent in two of the libraries produced (Rd sheared/pASC15 and Rd *Pvu*II/pASC15) but insertions into other sites in this same gene

were found in the other two Tn7 mutated libraries. The gain-of-function mutant in the Tn7 *in vitro* system has been reported to remove the sequence bias observed with the wild-type transposase (95) so the potential effect of user error, including contaminated reagents, cannot be ignored. Rather than attempting to further dissect this problem, the GPS mutagenesis system was abandoned for a Tn5-based *in vitro* mutagenesis system that had recently become available. The sites of mapped insertions using the GSP-1 Tn7 system are listed in Appendices A and B. The sites of Tn7 insertions into the pASC13 and pASC15 vector sequences are shown in Figure 2.5.

Tn5 mutagenesis with pASC15. The EZ::TN mutagenesis system from Epicentre Technologies was used to create the transposon insertions for the remainder of this work. It was first utilized to mutate the same pASC15 library containing *Pvu*II-digested Rd KW20 chromosomal DNA that was used above. A total of 161 plasmids were sequenced representing 165 total insertions. Insertions into *H. influenzae* sequences accounted for 96 of the mapped insertions and 92 of these were at unique sites. These insertions were found in 39 different genes and 3 intergenic regions. While many of these were clustered in adjacent genes, the complexity of this library was much greater than might have been suggested by the results from the Tn7 mutagenesis. A potential problem was an increase in the frequency of insertions recovered in vector sequences. A total of 69 insertions in the vector were sequenced at 44 unique vector sites, representing a 41.8% rate of vector insertions. This frequency was higher than the 12.1% rate observed with the Tn7 transposon using the same chromosomal library. The pattern of Tn5 insertions into pASC15 is shown in Figure 2.6. Analysis of the insertion patterns into the vector

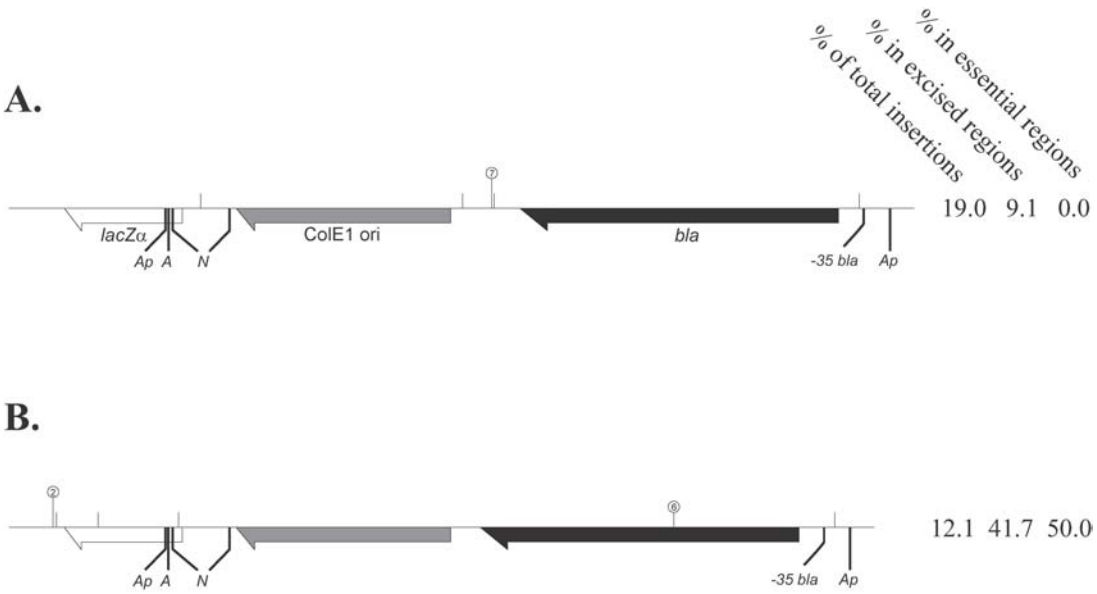


Figure 2.5 – Distribution of Tn7 insertions in pASC vectors. A-B. Locations of mapped Tn7 transposon insertions in pASC13 and pASC15 vectors, respectively. Numbers of insertions recovered at the same site are respresented by the figures in circles above the insertion site. Definitions for the fields to the right: *% of total insertions*: the frequency of mapped vector insertions in relation to the total number of mapped insertions in libraries constructed in that vector; *% in excised regions*: the frequency of vector insertions in regions of the vector removed during minimalization of the plasmid libraries; *% in essential regions*: the frequency of vector insertions in regions of the vector essential to plasmid maintenance.

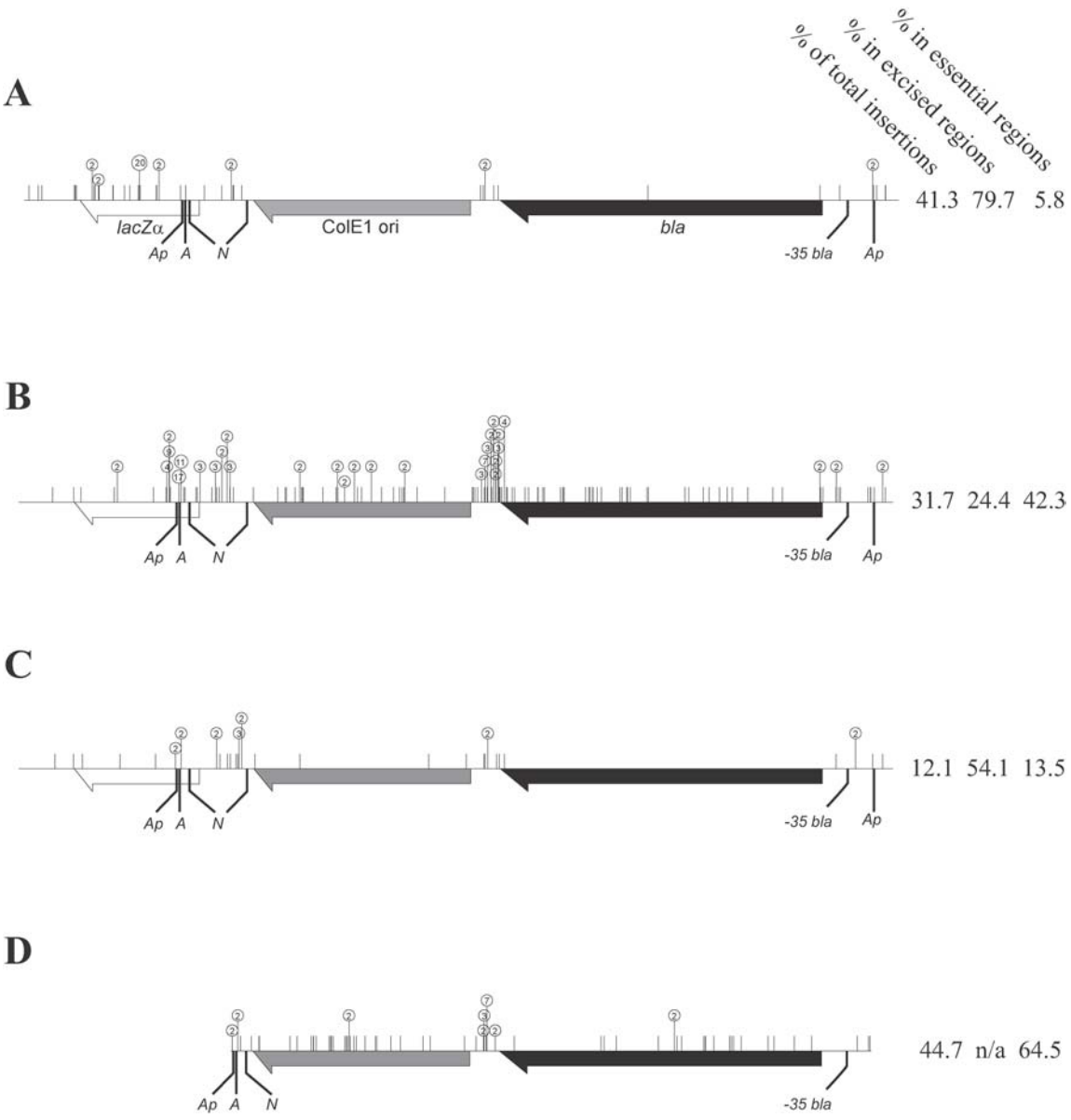


Figure 2.6 – Distribution of Tn5 insertions in pASC vectors. A. pASC15; **B.** pASC18, pre-mutagenesis minimalization; **C.** pASC18, post-mutagenesis minimalization; **D.** pASC18MIN**.** See Figure 2.5 for the definitions for the fields to the right.

indicates that 79.7% of the insertions were located within regions that should have been excised by the *Apa*I and *Not*I restriction digests; 5.8% were within essential regions of the vector (essential is defined as the origin of replication and *bla* coding region and associated promoter region). Insertions found in the excisable regions of the plasmid imply that the post-cloning minimalization was not complete. Insertions in the essential regions would imply either vector dimerization following the vector minimalization steps or permissive sites within those areas. The sites of the mapped Tn5 insertions in libraries using pASC15 are listed in Appendix C.

Tn5 transposition in pASC18; pre-mutagenesis minimalization. While the insertion sites in Tn5 appeared to be random, the clustering of insertions still indicated that the chromosomal libraries might not be sufficiently diverse even though the libraries appeared diverse when examined by agarose gel electrophoresis (data not shown). While the sizes of inserts and possible toxicity effects of certain genes may explain this problem, a potential issue that could reduce the complexity of the libraries is transcriptional interference with plasmid replication. In order to create the minimal vector, most of the sequence between the insertion site and replication start site were removed prior to transposon mutagenesis. If significant transcription occurs from promoters within the insert into the origin region of the plasmid, it might prevent replication or result in a lower copy number so that certain members of the plasmid library are underrepresented within the population. To alleviate this potential problem, a Rho-independent transcription termination sequence was inserted into pASC15 between the *Asc*I and *Not*I sites located within the *lacZα* gene. This terminator is retained in the plasmid after the post-cloning truncation of the vector backbone and is directionally

situated to prevent transcription from the insert into the origin region but does not result in termination of *lacZα* transcription. Chromosomal libraries were created in pASC18 using the restriction enzymes *Eco*RI, *Pvu*II, *Fsp*I, *Swa*I, and *Xmn*I and the plasmids were minimalized prior to Tn5 mutagenesis. A total of 573 plasmids were sequenced and 631 insert sites mapped. These sites included 432 inserts into chromosomal DNA (374 unique), including 96 insertions into rRNA operons (22.2%). The insertions were located in 127 CDSs and 13 intergenic sites. There were 199 insertions that mapped within the vector (123 unique), a rate of 31.5% of all mapped insertions (Figure 2.6). This represents a 23.3% decrease from the rate observed with pASC15. The dynamics of insertion site location in pASC18 was significantly different than that observed with pASC15. The rate of insertions mapped in excisable regions of the vector was 24.4% (49) and the rate in essential regions was 42.3% (85). This may be explained by more efficient removal of the excised vector sequences than in the previous library but an increased rate of recovery of vector dimers. The sites of mapped Tn5 insertion using the pre-mutagenesis minimalization protocol are listed in Appendix E.

Tn5 transposition in pASC18; post-mutagenesis minimalization. The high rate of recovery of transposon insertions in essential and excised vector sequences led to a change in the order of minimalization and mutagenesis. Rather than create a chromosomal library and excise the nonessential regions of the vector before transposon mutagenesis, we decided to create the library, mutate and then perform the excision steps. One benefit of this chronology is that recovery of insertions within essential regions should be much lower since, presumably, most of the dimers are the result of concatemers formed during the ligation steps that follow the *Apa*I and *Not*I digestions. A

sole library, created from *Psi*I digested Rd KW20 chromosomal DNA, was mutated using this protocol and 259 plasmids were sequenced. A total of 305 insertions were mapped with 268 found in chromosomal DNA sequences. Of these chromosomal inserts, 207 were in unique sites located in 54 CDSs and at 6 intergenic sites. Sixty-five of the 268 chromosomal insertions were mapped to rRNA operons (24.3%). Insertions into the vector sequence were recovered at a much lower frequency than in previous Tn5-mutated libraries. Of the 306 insertions, only 37 (28 unique) were located within vector sequence (12.1%). As was expected, considerably lower numbers were located within essential regions of the vector (13.5%). However, insertions continued to be recovered in regions that should have been excised during the truncation protocol (54.1%) (Figure 2.6).

Since plasmids with chromosomal inserts less than 2000-bp were removed, and the total length of the non-essential sequence in pASC18 is 198-bp, one would expect, in a random system, a maximum of 9.1% of the mapped insertions to be located in vector DNA. The frequency of recovery of Tn5 insertions into pASC18 using the post-mutagenesis minimalization method was 12.1%, which was not a significant increase over the expected value (P=0.284). However, there are two reasons why this statistical analysis may be inaccurate. First, the average size of the chromosomal inserts was larger than 2000-bp, so the expected frequency of vector insertions should actually be lower than 9.1%. Secondly, a large percentage of vector insertions (67.5%) were located in regions that should have been excised during post-mutagenesis minimalization or were in regions deemed essential to plasmid viability. The number of nucleotides in these regions was not accounted for when the expected number of vector insertions was determined. Regardless, the continued recovery of insertions into both non-essential and essential

regions of the vector indicated that a lower frequency of sequenced vector transposon insertions could still have been accomplished. The sites of mapped Tn5 insertions using the post-mutagenesis truncation protocol are listed in Appendix E.

Tn5 transposition in pASC18MIN. In a continued effort to lower the frequency of recovered vector insertions, a minimalized pASC variant was used that did not require post-cloning modification. Since the majority of sequenced vector inserts in the pASC15 and post-mutagenesis truncated (pASC18) libraries fell within areas of the vector that should have been removed after cloning, we decided to attempt to clone the chromosomal DNA directly into pASC18MIN, a variant of pASC18 in which the *lacZα* region had already been removed. This change leaves 182-bp (10.8%) of the vector that should be available for recovery of insertions. Because of the inability to screen plasmids for insertions, one major complication would be the likelihood of vector dimers being a prominent member of the mutagenesis pool. To lower the chances of vector dimers being present, the vector was treated with calf intestinal alkaline phosphatase (CIP) and phosphorylated adapters were used to clone chromosomal DNA into the *Asc*I site. The use of pASC18MIN, while faster and easier than the post-cloning truncation method, was unsuccessful in reducing vector insertions. We sequenced 152 plasmids with 168 mapped insertions (Appendix F). Of these, 74 were contained in the vector (61 unique) (Figure 2.6). This was a significant increase over the combined pASC18 libraries, 44.1% to 25.2%, and a dramatic increase over the 12.1% rate recovered from the post-mutation truncated pASC18 library. Examination of the sites of vector insertions indicated that incomplete dephosphorylation of the vector, leading to vector concatemers, is the likely reason for this increase; 64.5% of the insertions were recovered in essential regions of the

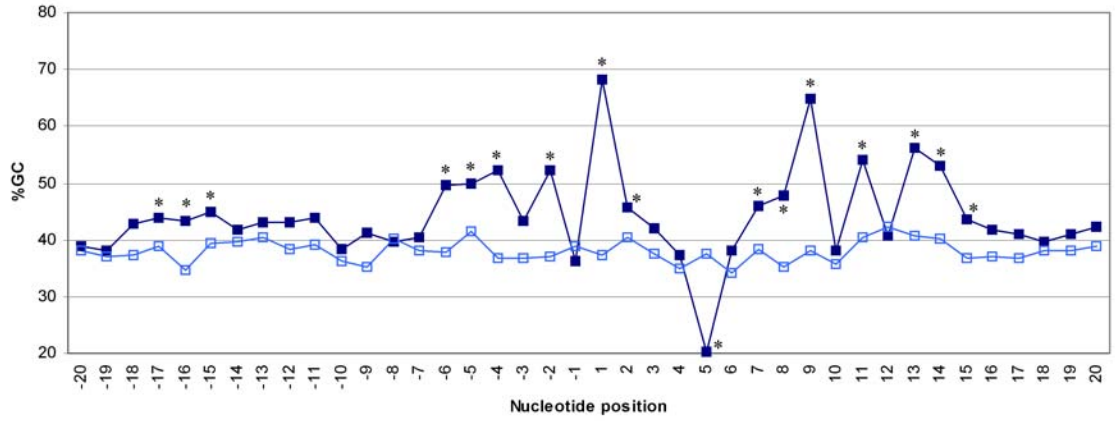
vector. The region between the end of *bla* and the start of the origin region accounted for 21.1% of the insertions. This region was to be a later target for additional deletions of the vector. Insertions into Rd KW20 chromosomal DNA accounted for 94 of the mapped insertions (57 unique, 27 rRNA) and were found in 31 CDSs and two intergenic sites. These results indicate that adapter-mediated cloning into pASC18MIN lowered the yield of chromosomal Tn5 insertions over that observed using the post-mutagenesis minimalization strategy.

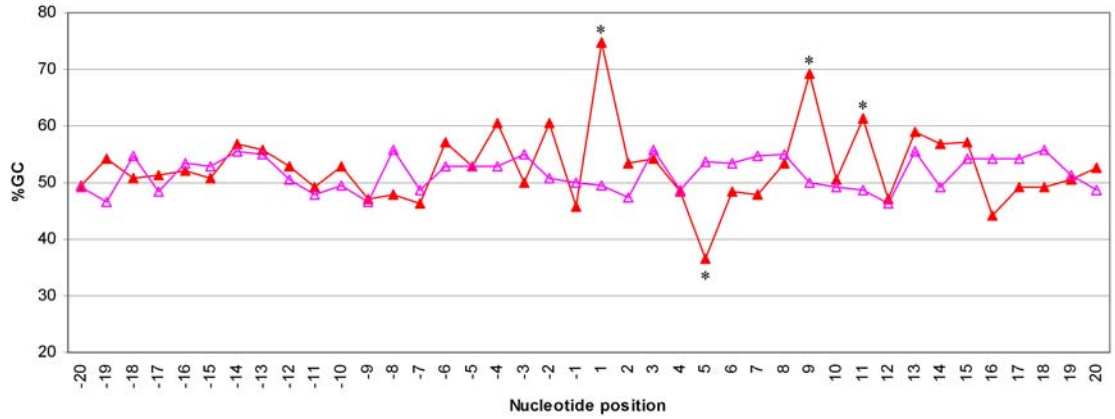
Analysis of Tn5 insertion site characteristics. During the Tn5 transposition event, the transposase cleaves the opposite strands of the target DNA at two sites 9-bp apart and integrates the transposon element. DNA repair at this site results in a 9-bp duplication with the intervening transposon sequence. For this reason, either of the nucleotides at position 1 or 9 (in reference to sequencing from one end of the element) could be considered the actual site of insertion. Sequencing of the Tn5 transposon insertion sites in this study was performed from one end of the element (using the KAN-2 RP1 primer). For ease of discussion, the nucleotide at the first position after the transposon mosaic end in the sequencing data is referred to as the insertion site.

The large number of Tn5 transposon insertions recovered in this study allowed for detailed analysis of potential biases in target site selection. Insertions into the vector and chromosomal fragments were separated and the 40 nucleotides immediately surrounding each transposon insertion site were analyzed for the frequency of single nucleotides and GC content at each position. The results of the analysis of the chromosomal insertions are shown in Figure 2.7 and Table 2.3. Positions of significant deviation from that of the randomly chosen 40-mers were determined using the Z-test method (95% confidence

level). Additionally, the overall GC content surrounding the Tn5 insertions were compared to the random sequences. A prominent feature of the insertion sites in this study was a higher than expected GC content in the immediate vicinity of the sites. The average GC content is 38.2% for the *H. influenzae* Rd KW20 genome and 38.0% for the 715 randomly chosen 40-mers used for comparison. In contrast, the average GC content of the 40 nucleotides surrounding the 715 Tn5 insertions mapped in *H. influenzae* sequences was 44.2%. Examination of GC content of the Tn5 insertions identifies a clear pattern of symmetry surrounding the fifth nucleotide from the mapped insertion site, and extends for 10 nucleotides on either side of this position. The average GC content of this 21-bp region in the Rd KW20 Tn5 insertion data set was 46.6% and was significantly higher than that expected by random chance (P=0.002). In contrast, the GC content of this same area around the 238 mapped vector insertions was 52.7% which was not significantly higher than the 51.6% observed in the random vector sequences (P=0.917).

Figure 2.7 – Average GC content surrounding the Tn5 insertion sites in Rd KW20 genomic and vector DNA sequences. The dark blue line (closed squares) indicates average GC content for the 40-bp surrounding the Tn5 insertions in Rd KW20 sequences (n=715). The light blue line (open squares) indicates average GC content of 715 randomly selected 40-mers from Rd KW20. The red line (closed triangles) indicates average GC content for the 40-bp surrounding the Tn5 vector insertions (n=238). The pink line (open triangles) indicates average GC content of 238 randomly selected 40-mers from pASC18. Nucleotide position one indicates the mapped site of insertion. Asterisks indicate positions within the Tn5 insertion sites of significant deviation from the random Rd or vector sequences ($P \leq 0.05$).


%GC
Nucleotide position


%GC
Nucleotide position

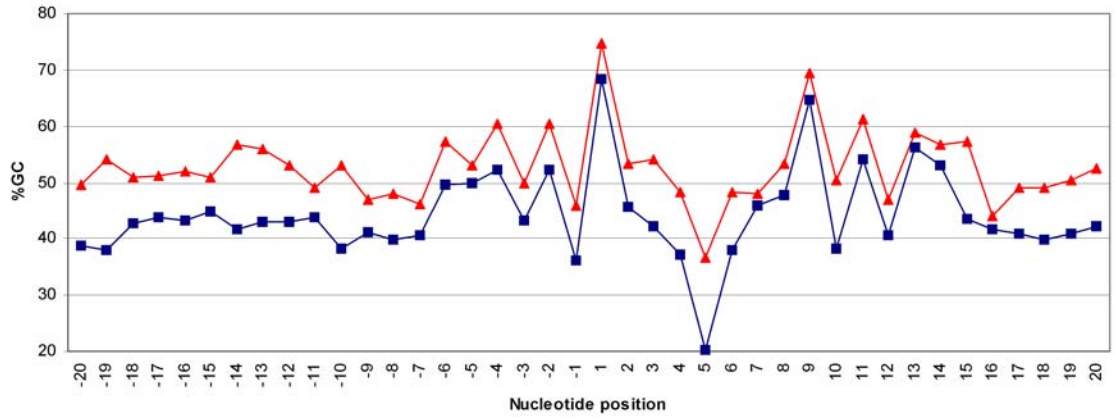

%GC
Nucleotide position

Table 2.3 – Frequency of nucleotides at positions surrounding the Tn5 insertion sites in Rd KW20 DNA

	-20	-19	-18	-17	-16	-15	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1
A	30.3	29.7	29.4	31.7	32.7	27.6	27.4	26.4	27.7	30.5	32.0	28.0	28.3	25.9	20.6	27.7	24.9	32.7	22.8	34.4
C	**23.4**	18.0	21.8	21.0	21.5	21.4	**23.6**	**23.4**	20.7	22.1	18.7	22.1	20.0	20.3	**27.7**	21.7	**30.2**	18.3	**37.1**	15.0
G	15.4	20.0	21.0	22.8	21.8	**23.5**	18.0	19.7	22.2	21.8	19.6	19.2	19.7	20.3	21.8	**28.1**	22.1	**25.0**	15.2	21.3
T	30.9	32.3	27.8	24.5	23.9	27.6	30.9	30.5	29.4	25.6	29.7	30.8	32.0	33.6	29.9	22.5	22.8	23.9	24.9	29.4
AT	61.3	62.0	57.2	56.2	56.6	55.1	58.3	56.9	57.1	56.1	61.7	58.7	60.3	59.4	50.5	50.2	47.7	56.6	47.7	63.8
GC	38.7	38.0	42.8	**43.8**	**43.4**	**44.9**	41.7	43.1	42.9	**43.9**	38.3	41.3	39.7	40.6	**49.5**	**49.8**	**52.3**	**43.4**	**52.3**	36.2

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
A	20.0	18.3	22.0	15.1	**40.0**	**46.2**	33.7	33.6	13.4	26.4	21.3	24.3	18.3	19.6	31.5	32.9	27.6	31.2	30.3	27.6
C	**25.5**	**28.5**	**24.9**	**27.7**	10.5	10.2	21.0	18.9	**41.0**	**26.3**	19.4	22.5	22.8	**35.1**	20.7	22.4	21.7	19.2	21.7	**23.6**
G	**42.8**	17.1	17.2	9.5	9.8	**27.8**	**25.0**	**28.8**	**23.8**	11.9	**34.5**	18.2	**33.4**	17.9	22.8	19.3	19.2	20.6	19.2	18.6
T	11.7	36.1	35.9	**47.7**	**39.7**	15.8	20.3	18.7	21.8	35.4	24.8	35.0	25.5	27.4	25.0	25.5	31.6	29.1	28.8	30.2
AT	31.7	54.4	57.9	62.8	**79.7**	62.0	54.0	52.3	35.2	61.8	46.0	59.3	43.8	47.0	56.5	58.3	59.2	60.3	59.2	57.8
GC	**68.3**	**45.6**	42.1	37.2	20.3	38.0	**46.0**	**47.7**	**64.8**	38.2	**54.0**	40.7	**56.2**	**53.0**	**43.5**	41.7	40.8	39.7	40.8	42.2

Nucleotide position 1 indicates the mapped site of insertion.
Bold or red numbers indicate a significantly higher or lower frequency, respectively, than expected ($P \leq 0.05$)

The higher than expected GC content observed at the Rd KW20 insertion sites implies that a bias existed. There are two possible explanations for this observation. The targeting of insertion sites could be biased by sequence context either immediately surrounding the site or a much larger region around the site. Alternatively, the chromosomal libraries themselves could have been biased and the higher GC content is simply a reflection of this fact.

Shevchenko *et al.* used the Tn5 *in vitro* system to introduce priming islands for systematic sequencing of cDNA clones (90). They examined the sites of more than 24,000 Tn5 insertions into 1955 cDNA clones. The distribution of insertions into clones was not completely random but GC content did not seem to explain the biases. Oddly, they did not report what species the cDNA represented but from the graphs presented it appeared that the sequences averaged around 54-55% G+C. The 21-bp region surrounding the duplication site averaged 54.0% G+C. The major disparity between the AT-rich sequence of the *H. influenzae* genome and their GC-rich cDNA clones may account for the different observations of a GC bias.

Global Tn5 insertion profiles in *H. influenzae* Rd KW20 were examined to see if transposon insertions tended to be recovered in areas of higher GC content. There are 6 rRNA operons present in the Rd KW20 genome and they account for 1.8% of the total genomic content. In this work, unique Tn5 insertions into the rRNA operons accounted for 165 of the 736 unique insertions into cloned genomic DNA (22.4%). This could indicate that rRNA DNA was over-represented in the libraries; however, the average GC content of the Rd KW20 rRNA operons is 48.0% compared to the 38.2% observed for the whole of the genomic sequence. Similarly, examination of the GC content of the Rd

KW20 annotated CDSs and mapped Tn5 insertions in this study indicates a statistically significant bias towards Tn5 insertions into CDSs with higher GC content (P=0.001). In Table 2.4, Rd KW20 CDSs were grouped with respect to GC content deviation from the mean. The expected number of CDSs hit in each group was calculated based on n = 217 (the actual number of CDSs hit) and compared to the observed results. No insertions were recovered in any CDS with a GC content less than two standard deviations (SD) from the mean. In contrast, the relative rate of Tn5 insertion into each group increased as GC content increased. Oddly, most of the Tn5 sites in the most GC-poor genes (greater than 1 standard deviation below the average %GC) were located near the junction with the vector sequence (average 149 bp from junction).

The average length of CDSs hit (1357.7 bp) was greater than the average length of CDSs in the Rd KW20 genome (917.9 bp). This observation might be expected considering that the longer the CDS the greater likelihood that at least a portion of the sequence would be present in a plasmid mutagenesis library and that it would be subject to Tn5 insertion. Therefore, one alternate explanation for the apparent GC content bias is that the size of the CDSs within each group increased making Tn5 insertions more likely. For all Rd KW20 CDSs, the size and GC content demonstrated only a small level of correlation (0.165, P=0.00). To further examine the possibility of gene size biasing the distribution of insertions into the GC SD groups, the expected number of insertions into each group was determined based upon the total size of the CDSs within that group and compared to the observed number (Table 2.4). This examination indicated that the observed GC content bias could not be accounted for by the relative sizes of the CDSs within each group.

Table 2.4 – Distribution of Tn5 insertions into Rd KW20 CDS in relation to GC content

		Rd KW20		Tn5 Rd insertions [a]					
				Observed		By %GC content		By length[b]	
SD [c]	%GC	CDS	Frequency	CDS	Frequency	Expected [d]	Relative rate [e]	Expected [f]	Relative rate [e]
> -3SD	< 27.66	15	0.88	0	0.00	1.90	0.00	1.27	0.00
-2SD to -3SD	27.67-31.26	44	2.57	0	0.00	5.58	0.00	3.73	0.00
-1SD to -2SD	31.27-34.86	171	10.00	8	3.69	21.70	0.37	15.50	0.52
Mean to -1SD	34.87-38.46	567	33.16	44	20.28	71.95	0.61	67.33	0.65
Mean to +1SD	38.47-42.07	701	40.99	116	53.46	88.96	1.30	101.89	1.14
+1SD to +2SD	42.08-45.67	175	10.23	35	16.13	22.21	1.58	21.41	1.63
+2SD to +3SD	45.68-49.27	30	1.75	12	5.53	3.81	3.15	3.52	3.41
> +3SD	>49.28	7	0.41	2	0.92	0.89	2.25	0.38	5.30

[a] Represents 217 CDS recovered with Tn5 insertions. Does not include multiple insertions into the same gene or insertions into intergenic/rRNA sequences.

[b] In relation to total length of CDS within each GC content standard deviation.

[c] Standard deviations from the mean GC content. The range of GC content is listed to the right.

[d] Number of CDS expected with respect to GC content (n=217).

[e] CDS recovered divided by CDS expected.

[f] Number of CDS expected with respect to length. Total length (bp) of CDS within GC range divided by total length of all CDS x 217.

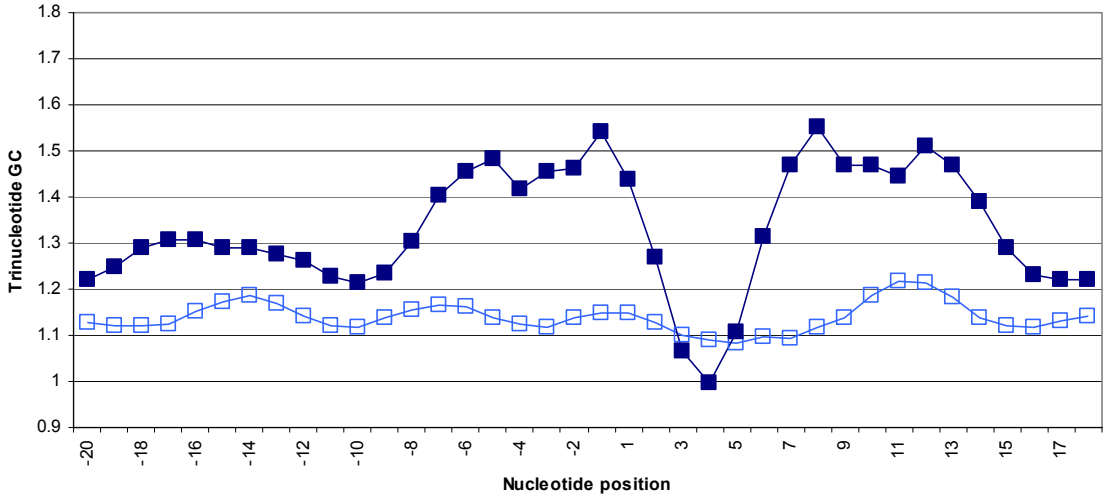
An alternate explanation for the higher GC content observed in the sequence immediately surrounding the insertion site and the tendency to recover insertions in rRNA operons and in CDSs with higher GC content may be that the chromosomal libraries themselves were biased. While the GC bias was seen for insertions in all of the mutated libraries, the restriction enzymes used to create them varied greatly in their recognition sequences. The four major restriction enzymes used to create the libraries had the following recognition sequences: *Psi*I 5'-TTATAA; *Pvu*II 5'-CGATCG; *Swa*I 5'-ATTTAAAT; *Xmn*I 5'-GAANNNNTTC. If these enzymes cut either too frequently or not frequently enough in regions of low GC content, then restriction fragments from these areas might not be found in the mutated libraries. In order to determine whether a bias due to these enzymes could have existed, the restriction patterns generated by these four enzymes were examined for the regions of the Rd KW20 genome that contain multiple CDSs with lower than average GC content. Because the libraries were also manipulated with *Asc*I, *Apa*I and *Not*I, the locations of these sites were also factored in. This analysis indicated that many regions of lower GC content should have been present in the libraries; in fact, some of these regions were present in the libraries but transposon insertions had not been recovered in the lower GC containing CDSs. Therefore, if a bias did occur within the libraries, the choices of restriction enzymes utilized in their construction would not have been the cause.

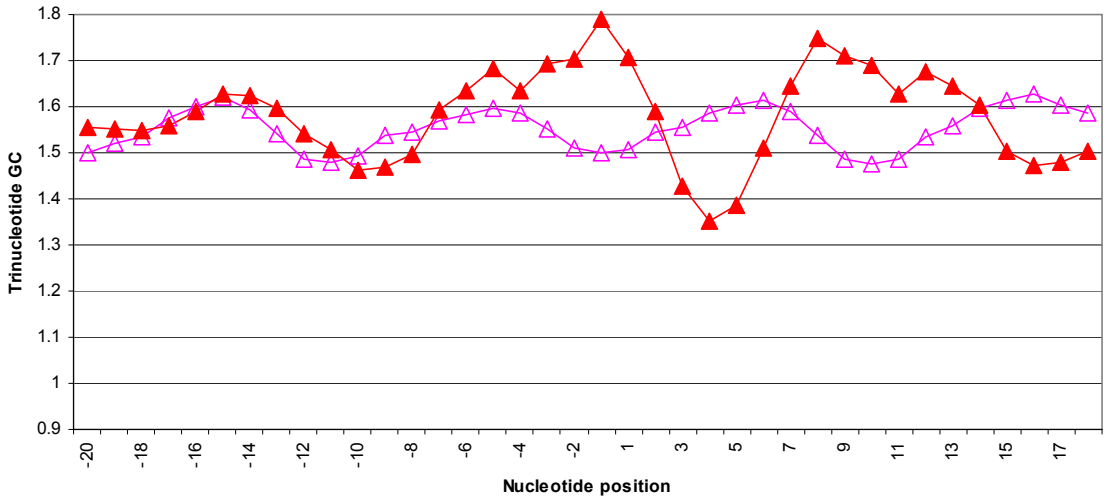
We had no independent data that established the randomness of restriction fragments libraries prior to mutagenesis. Therefore, it is impossible to conclusively state that the libraries themselves were not biased towards fragments of higher GC content.

Nevertheless, the increased insertion frequencies into CDSs as GC content increased implies that some sequence context affects the randomness of Tn5 insertion.

The work of Liao *et al.* and Vignal *et al.* with other transposable elements indicated that properties other than the explicit sequence at insertion sites, such as DNA secondary structure, might play a role in target site selection (59, 104). Liao *et al.* calculated that the DNA physical properties of A-philicity, protein-induced deformability (PID), B-DNA twist, and bendability were all shown to have a moderate to low correlation with dinucleotide or trinucleotide GC content (-0.413, 0.416, -0.135, and 0.097, respectively). Graphs indicating the average values of trinucleotide GC content and each of the above properties at positions surrounding the Tn5 insertion sites are shown in Figures 2.8 to 2.12 and compared to values from the random 40-mer sequences from Rd KW20 and pASC18. Analysis of trinucleotide GC content in *H. influenzae* and vector transposon insertions (Figure 2.8) indicates a striking resemblance of the signal generated for both, although the vector sequence seems shifted to a slightly higher GC content. There appears to be a qualitative difference between the trinucleotide GC content of the chromosomal insertions and the randomly selected Rd KW20 sequences when compared to the differences between the vector insertions and random sequences. This same phenomenon is apparent when examining patterns of A-philicity, PID, B-DNA twist and bendability surrounding the chromosomal and vector insertions. As noted with both the P transposable element and *mariner*-family of transposons, the lack of a clearly defined consensus sequences combined with significant signals apparent around the insertion sites when examining various DNA properties suggests that Tn5 recognizes some aspect of DNA structure rather than a specific DNA sequence.

Figure 2.8 – Average trinucleotide GC content surrounding the Tn5 insertion sites in *H. influenzae* Rd KW20 genomic DNA and vector sequences. The dark blue line (closed squares) represents the average values of the 40-bp surrounding the Tn5 insertions into Rd DNA (n=715) and the light blue line (open squares) represents the average values derived from 715 randomly selected 40-mers from Rd KW20. The red line (closed triangles) represents the average values of the 40-bp surrounding the Tn5 insertions into the vector (n=238) and the pink line (open triangles) represents the average values derived from 238 randomly selected 40-mers from pASC18. When multiple insertions were recovered at the same site, only one was included in the calculations unless evidence of independent insertions was available. Position 1 indicates the mapped site of insertion.


Trinucleotide GC
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1
0.9
-20 -18 -16 -14 -12 -10 -8 -6 -4 -2 1 3 5 7 9 11 13 15 17
Nucleotide position


Trinucleotide GC
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1
0.9
-20 -18 -16 -14 -12 -10 -8 -6 -4 -2 1 3 5 7 9 11 13 15 17
Nucleotide position

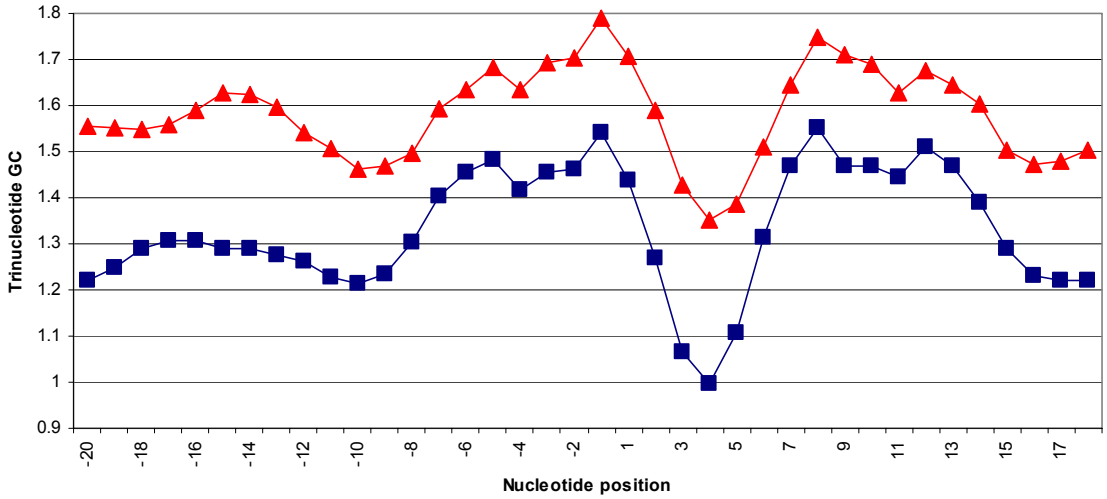
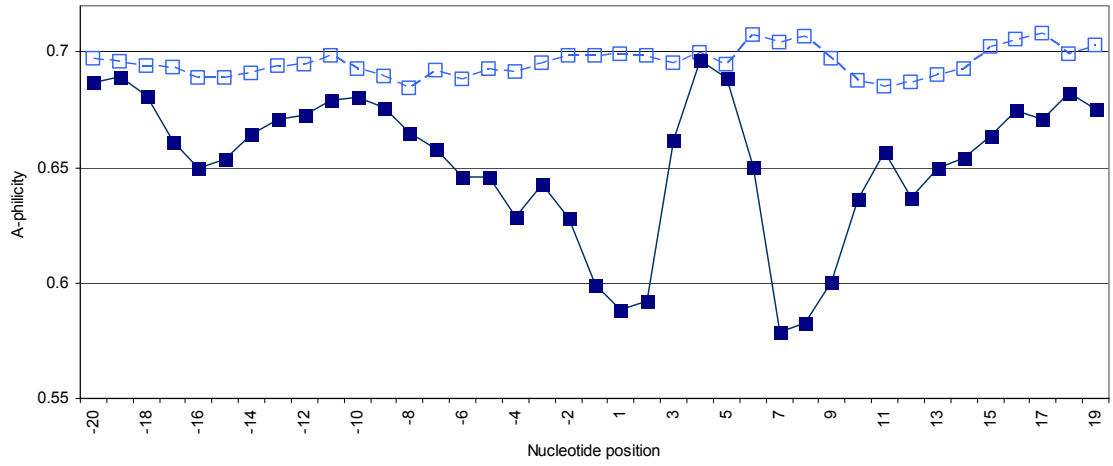
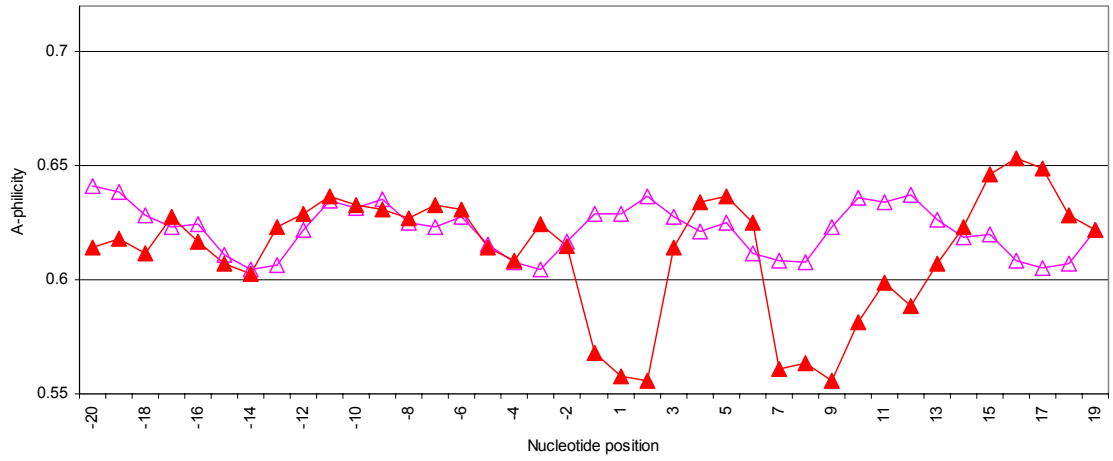

Trinucleotide GC
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1
0.9
-20 -18 -16 -14 -12 -10 -8 -6 -4 -2 1 3 5 7 9 11 13 15 17
Nucleotide position

Figure 2.9 – Average A-philicity surrounding the Tn5 insertion sites in *H. influenzae* Rd KW20 genomic DNA and vector sequences. The dark blue line (closed squares) represents the average values of the 40-bp surrounding the Tn5 insertions into Rd DNA (n=715) and the light blue line (open squares) represents the average values derived from 715 randomly selected 40-mers from Rd KW20. The red line (closed triangles) represents the average values of the 40-bp surrounding the Tn5 insertions into the vector (n=238) and the pink line (open triangles) represents the average values derived from 238 randomly selected 40-mers from pASC18. When multiple insertions were recovered at the same site, only one was included in the calculations unless evidence of independent insertions was available. Position 1 indicates the mapped site of insertion.


A-philicity
0.7
0.65
0.6
0.55
-20 -18 -16 -14 -12 -10 -8 -6 -4 -2 1 3 5 7 9 11 13 15 17 19
Nucleotide position


A-philicity
0.7
0.65
0.6
0.55
-20 -18 -16 -14 -12 -10 -8 -6 -4 -2 1 3 5 7 9 11 13 15 17 19
Nucleotide position

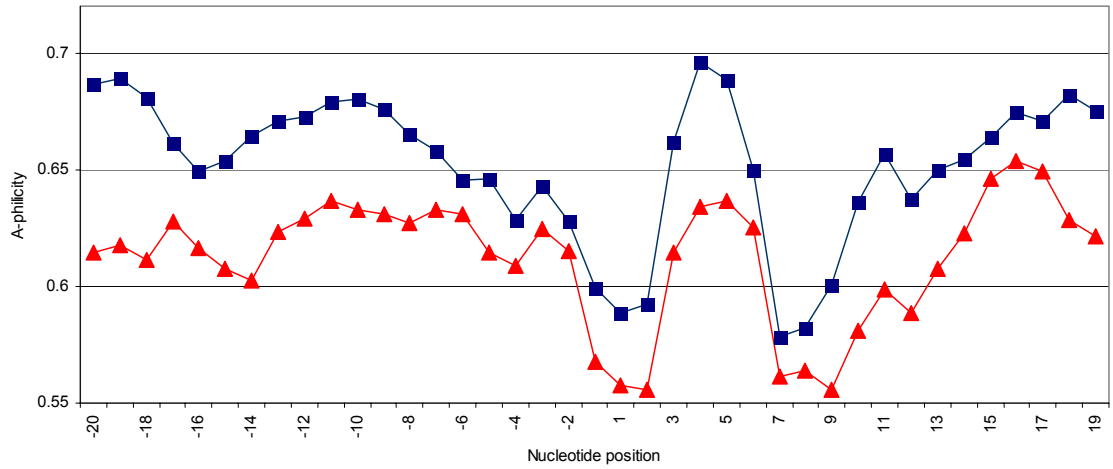
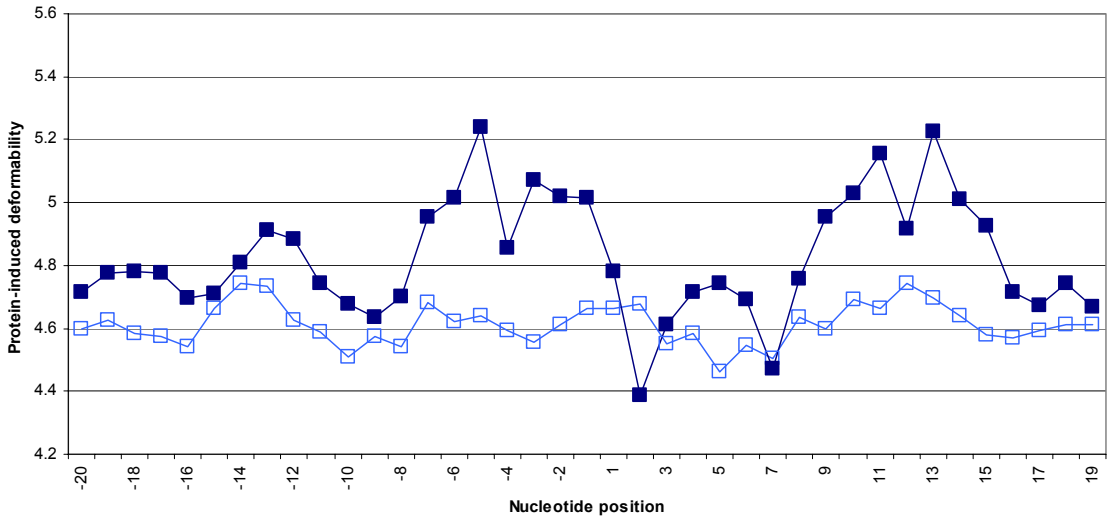
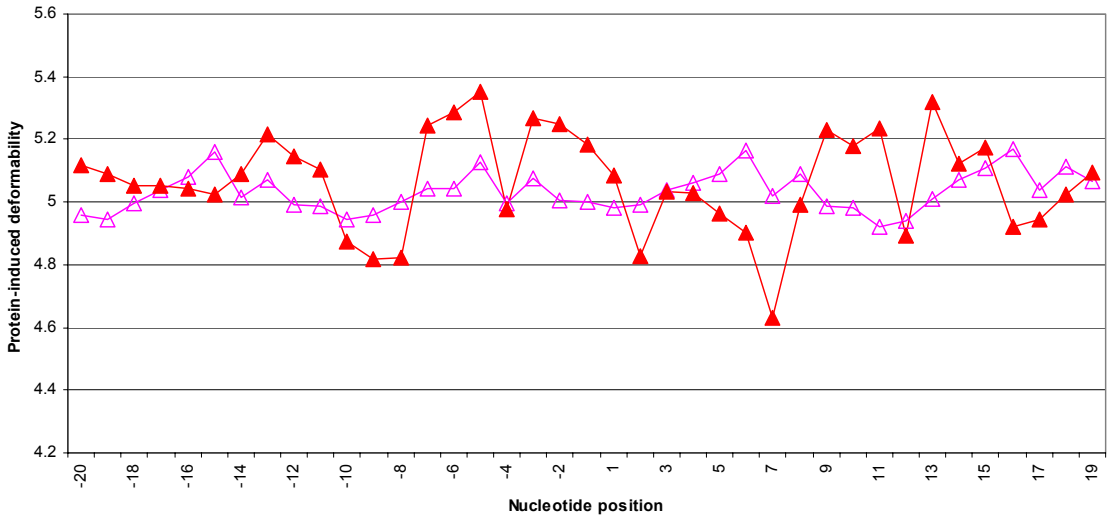

A-philicity
0.7
0.65
0.6
0.55
-20 -18 -16 -14 -12 -10 -8 -6 -4 -2 1 3 5 7 9 11 13 15 17 19
Nucleotide position

Figure 2.10 – Average protein-induced deformability surrounding the Tn5 insertion sites in *H. influenzae* Rd KW20 genomic DNA and vector sequences. The dark blue line (closed squares) represents the average values of the 40-bp surrounding the Tn5 insertions into Rd DNA (n=715) and the light blue line (open squares) represents the average values derived from 715 randomly selected 40-mers from Rd KW20. The red line (closed triangles) represents the average values of the 40-bp surrounding the Tn5 insertions into the vector (n=238) and the pink line (open triangles) represents the average values derived from 238 randomly selected 40-mers from pASC18. When multiple insertions were recovered at the same site, only one was included in the calculations unless evidence of independent insertions was available. Position 1 indicates the mapped site of insertion.


Protein-induced deformability
5.6
5.4
5.2
5
4.8
4.6
4.4
4.2
-20 -18 -16 -14 -12 -10 -8 -6 -4 -2 1 3 5 7 9 11 13 15 17 19
Nucleotide position


Protein-induced deformability
5.6
5.4
5.2
5
4.8
4.6
4.4
4.2
-20 -18 -16 -14 -12 -10 -8 -6 -4 -2 1 3 5 7 9 11 13 15 17 19
Nucleotide position

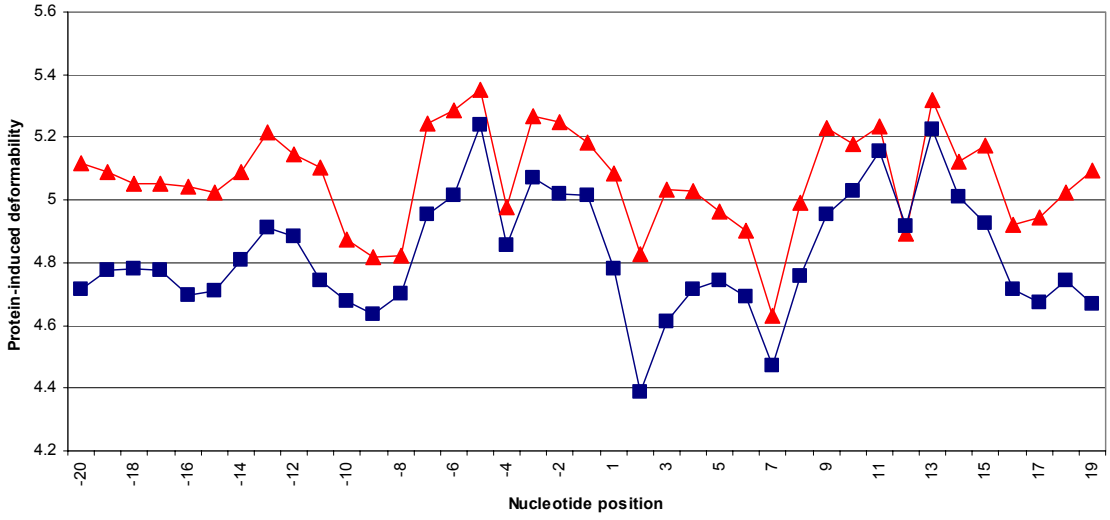
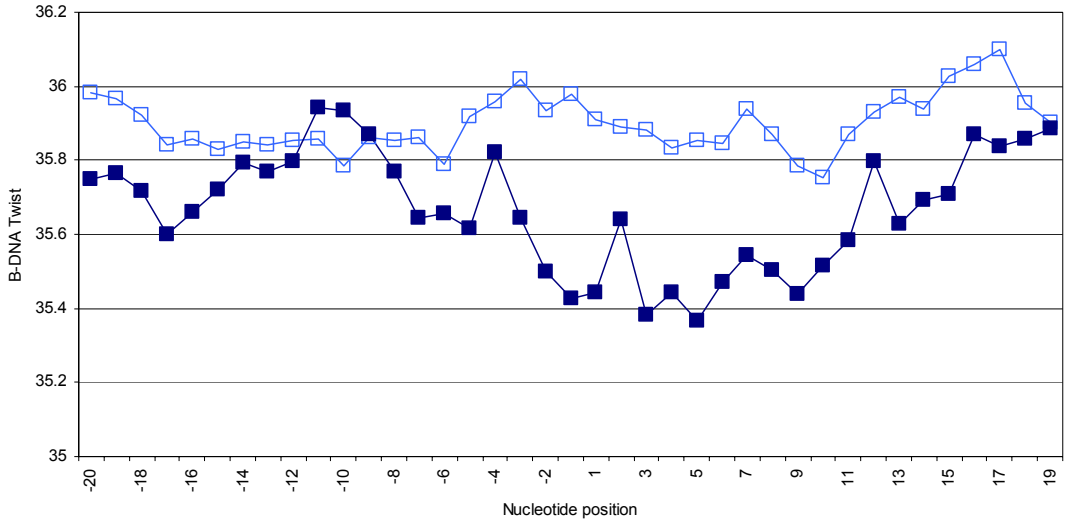
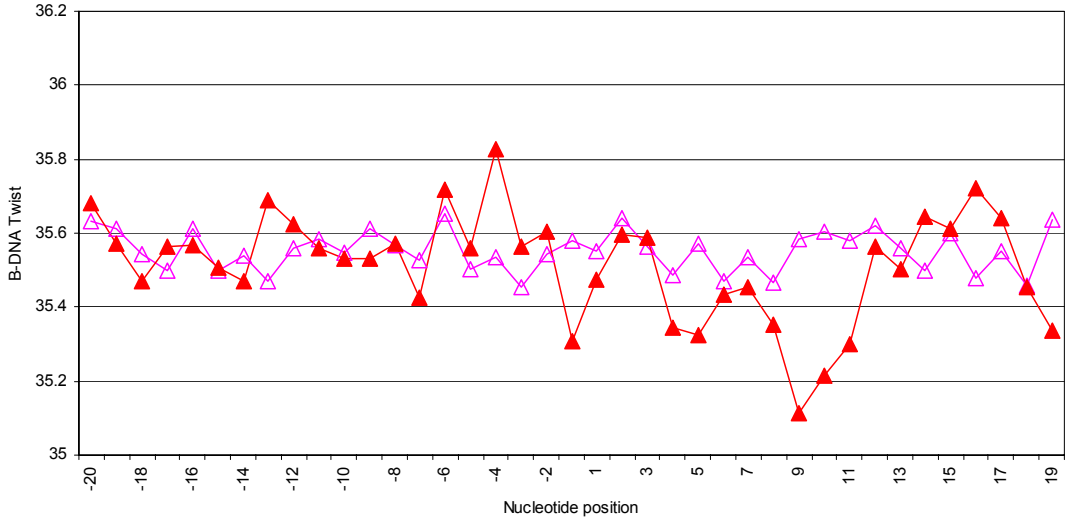

Protein-induced deformability
5.6
5.4
5.2
5
4.8
4.6
4.4
4.2
-20 -18 -16 -14 -12 -10 -8 -6 -4 -2 1 3 5 7 9 11 13 15 17 19
Nucleotide position

Figure 2.11 – Average B-DNA twist surrounding the Tn5 insertion sites in *H. influenzae* Rd KW20 genomic DNA and vector sequences. The dark blue line (closed squares) represents the average values of the 40-bp surrounding the Tn5 insertions into Rd DNA (n=715) and the light blue line (open squares) represents the average values derived from 715 randomly selected 40-mers from Rd KW20. The red line (closed triangles) represents the average values of the 40-bp surrounding the Tn5 insertions into the vector (n=238) and the pink line (open triangles) represents the average values derived from 238 randomly selected 40-mers from pASC18. When multiple insertions were recovered at the same site, only one was included in the calculations unless evidence of independent insertions was available. Position 1 indicates the mapped site of insertion.


36.2
36
35.8
35.6
35.4
35.2
35
B-DNA Twist
-20 -18 -16 -14 -12 -10 -8 -6 -4 -2 1 3 5 7 9 11 13 15 17 19
Nucleotide position


36.2
36
35.8
35.6
35.4
35.2
35
B-DNA Twist
-20 -18 -16 -14 -12 -10 -8 -6 -4 -2 1 3 5 7 9 11 13 15 17 19
Nucleotide position

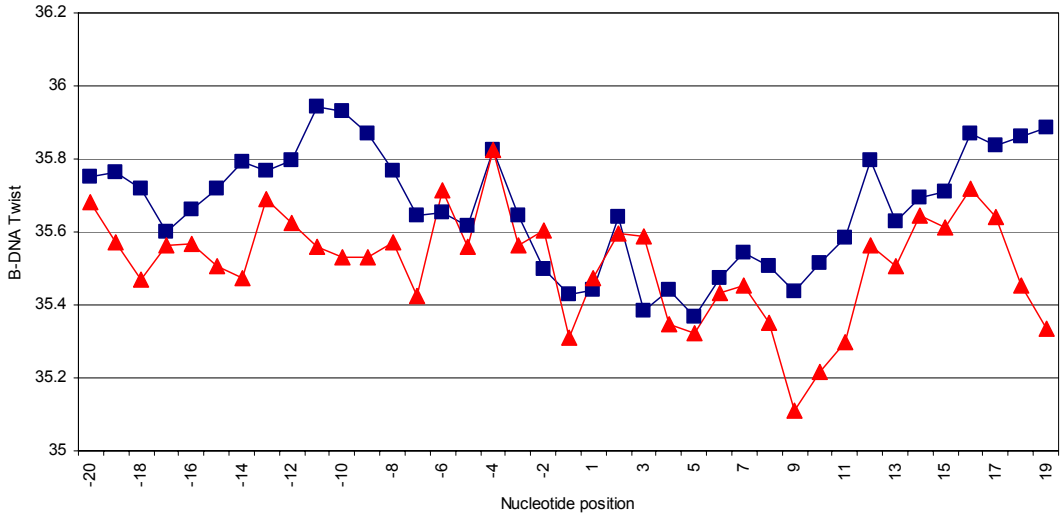
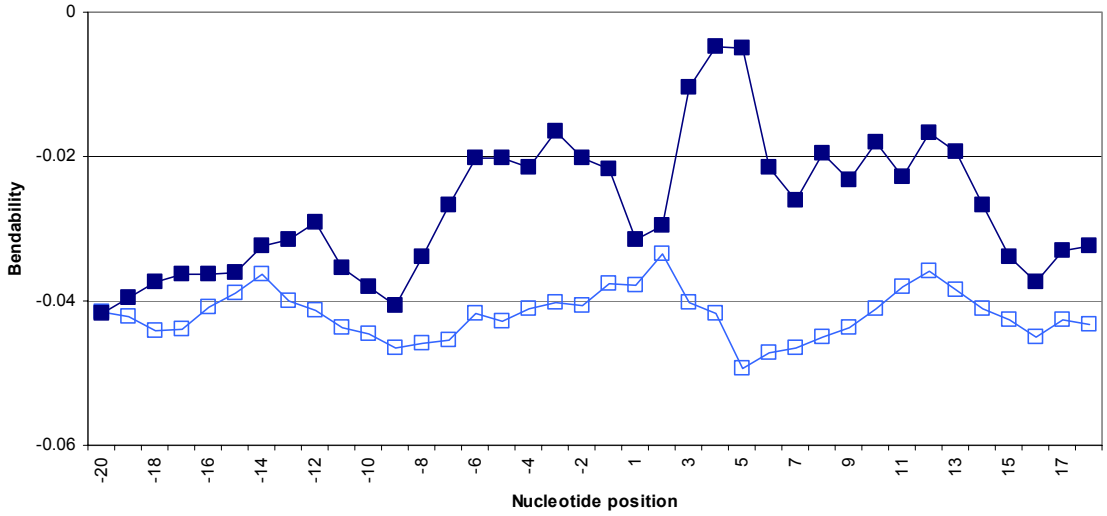
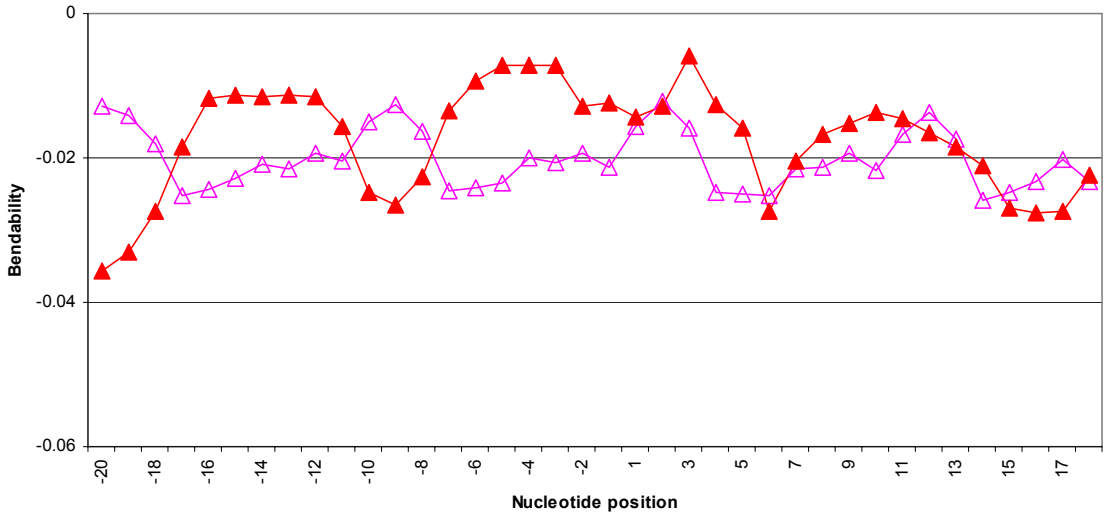
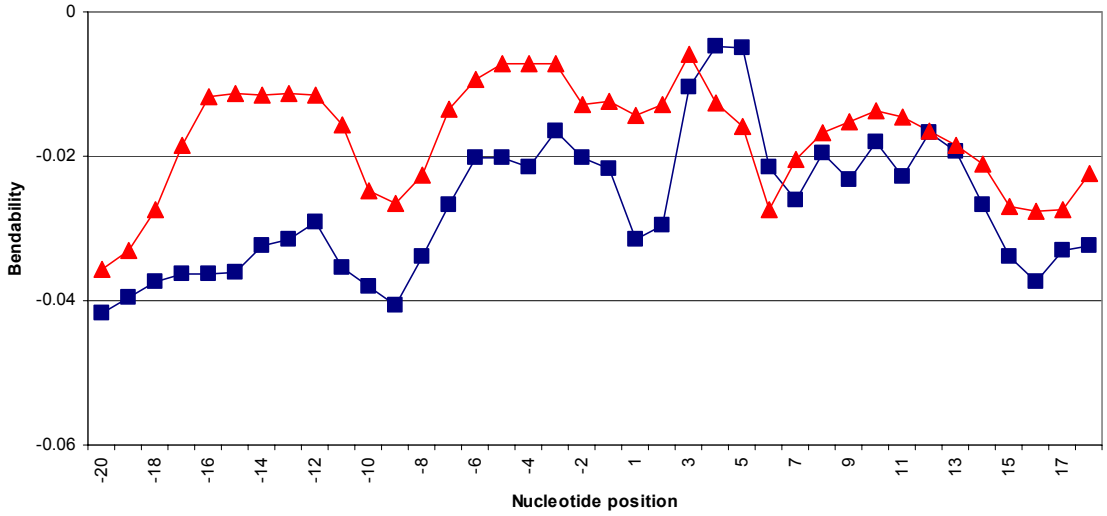

36.2
36
35.8
35.6
35.4
35.2
35
B-DNA Twist
-20 -18 -16 -14 -12 -10 -8 -6 -4 -2 1 3 5 7 9 11 13 15 17 19
Nucleotide position

Figure 2.12 – Average bendability surrounding the Tn5 insertion sites in *H. influenzae* Rd KW20 genomic DNA and vector sequences. The dark blue line (closed squares) represents the average values of the 40-bp surrounding the Tn5 insertions into Rd DNA (n=715) and the light blue line (open squares) represents the average values derived from 715 randomly selected 40-mers from Rd KW20. The red line (closed triangles) represents the average values of the 40-bp surrounding the Tn5 insertions into the vector (n=238) and the pink line (open triangles) represents the average values derived from 238 randomly selected 40-mers from pASC18. When multiple insertions were recovered at the same site, only one was included in the calculations unless evidence of independent insertions was available. Position 1 indicates the mapped site of insertion.


0
-0.02
-0.04
-0.06
Bendability
-20 -18 -16 -14 -12 -10 -8 -6 -4 -2 1 3 5 7 9 11 13 15 17
Nucleotide position

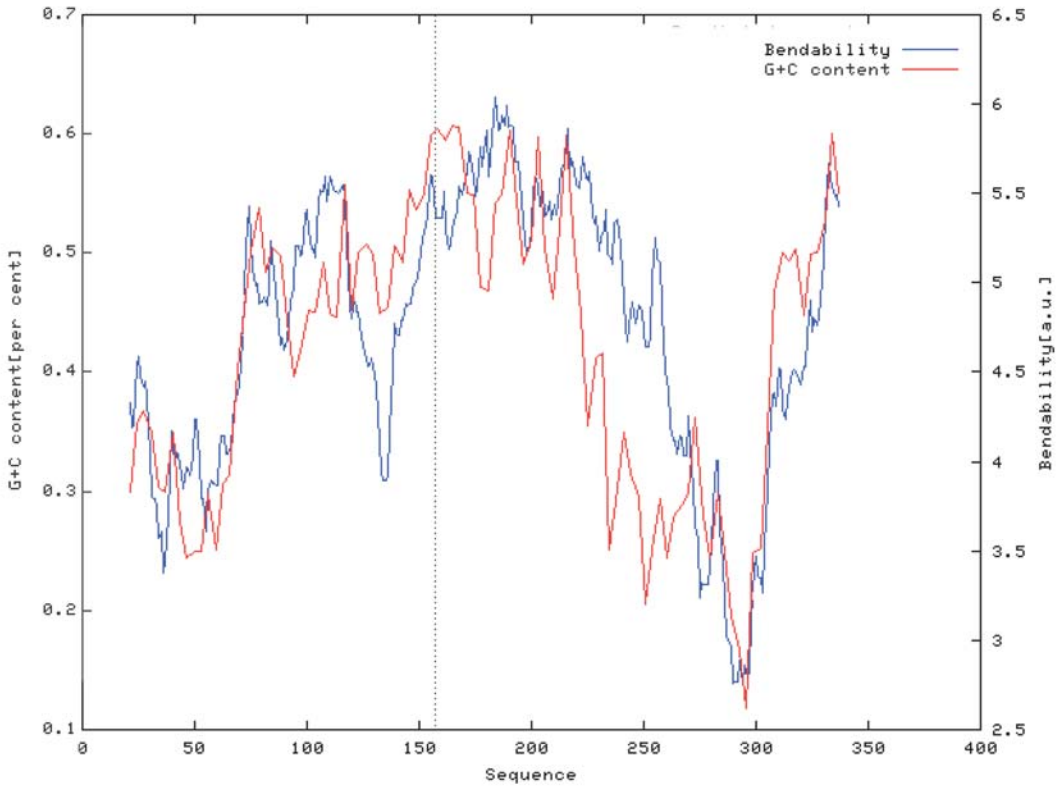

0
-0.02
-0.04
-0.06
Bendability
-20 -18 -16 -14 -12 -10 -8 -6 -4 -2 1 3 5 7 9 11 13 15 17
Nucleotide position


0
-0.02
-0.04
-0.06
Bendability
-20 -18 -16 -14 -12 -10 -8 -6 -4 -2 1 3 5 7 9 11 13 15 17
Nucleotide position

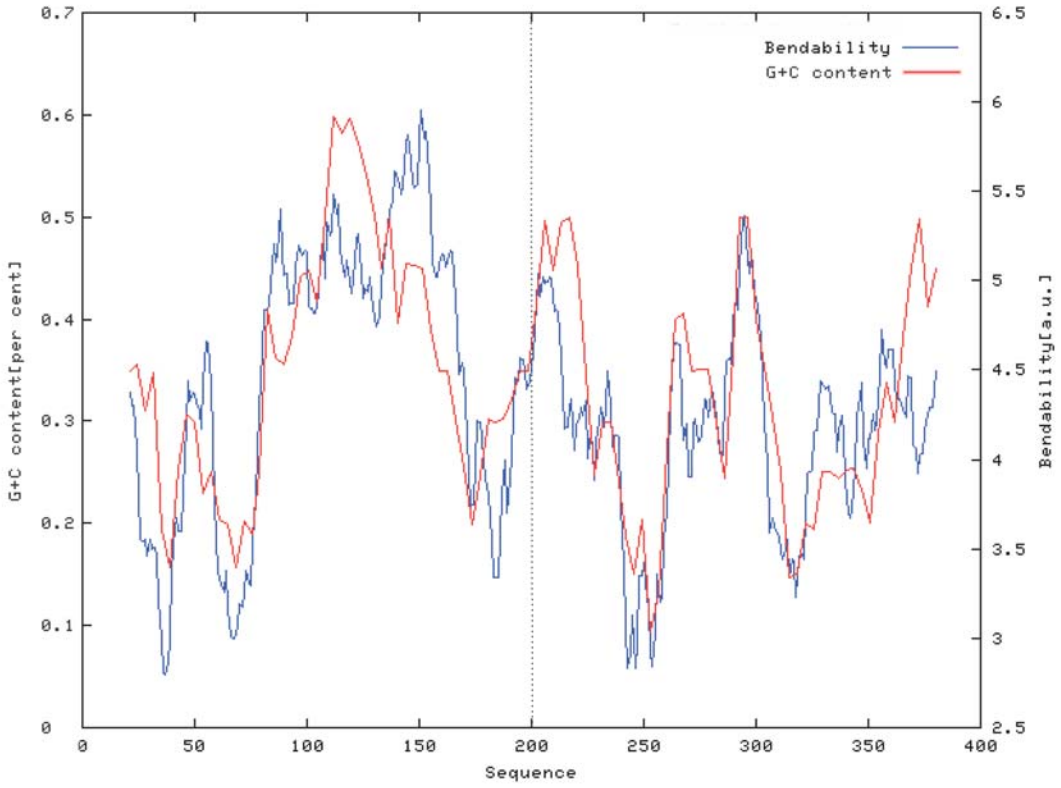
Of these properties, bendability might have the most biological significance (see discussion). Using the Bend.it program, the GC content and propensity for bendability were plotted for an approximately 400-bp region around the Tn5 insertion sites in GC-poor genes. These properties were also plotted for regions in which multiple Tn5 insertions were recovered in a short span of nucleotides. Plots from representatives of these two groups are shown in Figures 2.13 and 2.14, respectively. The sites of the Tn5 insertions in the rRNA operons were also plotted (Figure 2.15). While it was not possible to match the insertion sites recovered in rRNA to a specific operon, the six Rd KW20 rRNA operons are very homogenous. The only differences between the operons exist in the sequences between the 16S and 23S subunits, an area in which only a single Tn5 insertion was recovered. Analysis of the 21-bp surrounding the center of the Tn5 insertions indicated that while bendability was not at its maximum at each insertion site, the insertions tended to be located in peaks of increased bendability and increased GC content. Insertions were rare in areas characterized by prolonged decreases in bendability, such as those occurring in intergenic regions.

Figure 2.13 – Bendability and GC content plots for several Rd KW20 low G+C CDS with recovered Tn5 insertions. **A.** HI0661 (33.1% G+C); **B.** HI0973 (33.6% G+C); **C.** HI0588 (33.2% G+C); **D.** HI0216 (34.9% G+C). A 21-bp window was used to estimate bendability and GC content at each nucleotide surrounding the Tn5 insertion site (blue and red lines respectively). The dashed vertical lines represent the center of the 9-bp repeat generated by the Tn5 insertion.

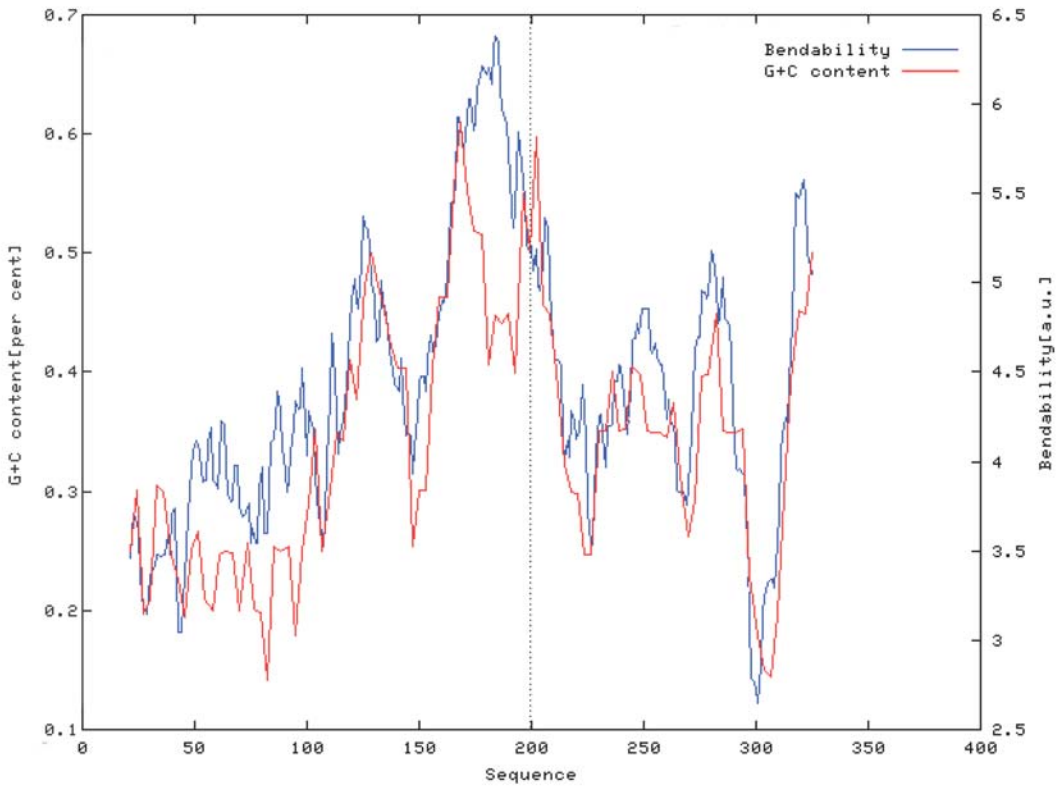
A


Bendability
G+C content
G+C content[per cent]
Bendability[a.u.]
Sequence
0.1
0.2
0.3
0.4
0.5
0.6
0.7
2.5
3
3.5
4
4.5
5
5.5
6
6.5
0
50
100
150
200
250
300
350
400

B


Bendability
G+C content
G+C content[per cent]
Bendability[a.u.]
Sequence
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
2.5
3
3.5
4
4.5
5
5.5
6
6.5
0
50
100
150
200
250
300
350
400

C


Bendability
G+C content
G+C content[per cent]
Bendability[a.u.]
Sequence
0.1
0.2
0.3
0.4
0.5
0.6
0.7
2.5
3
3.5
4
4.5
5
5.5
6
6.5
0
50
100
150
200
250
300
350
400

D

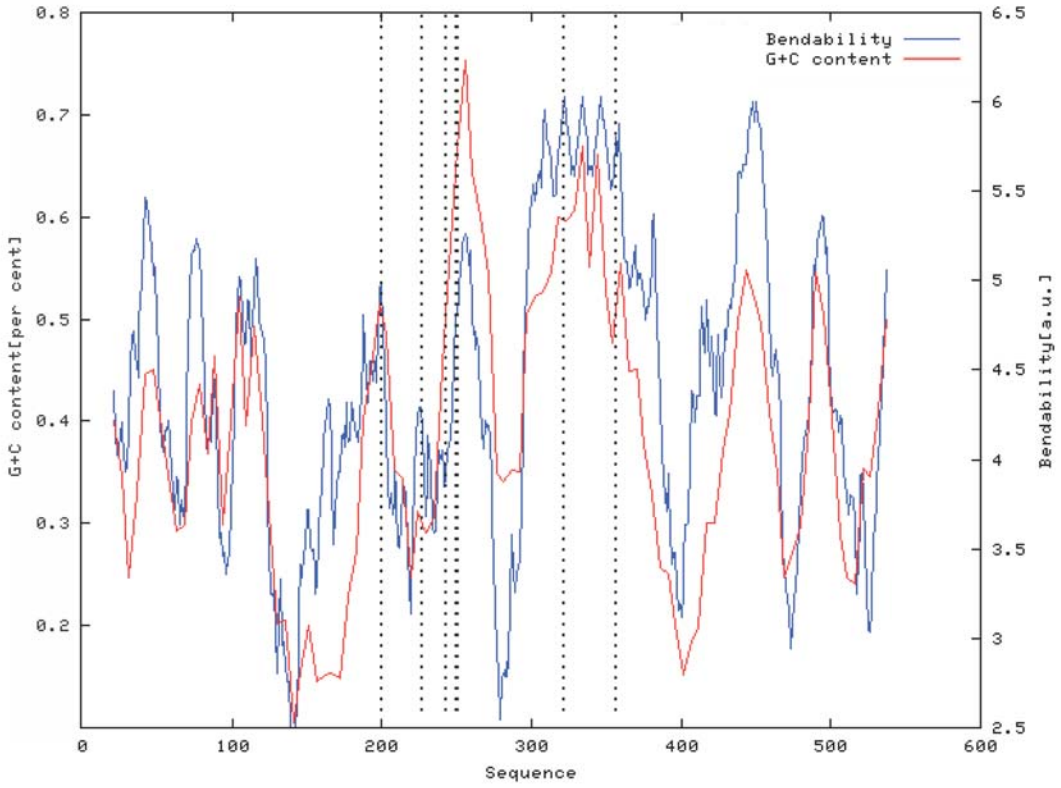
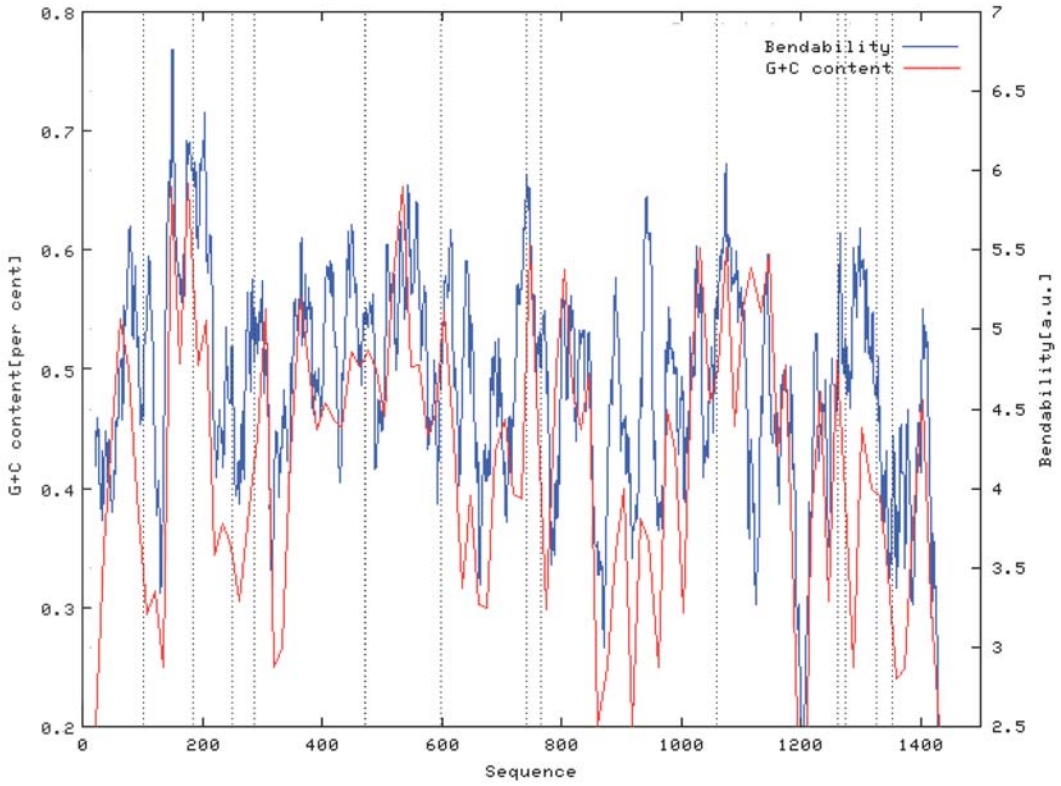
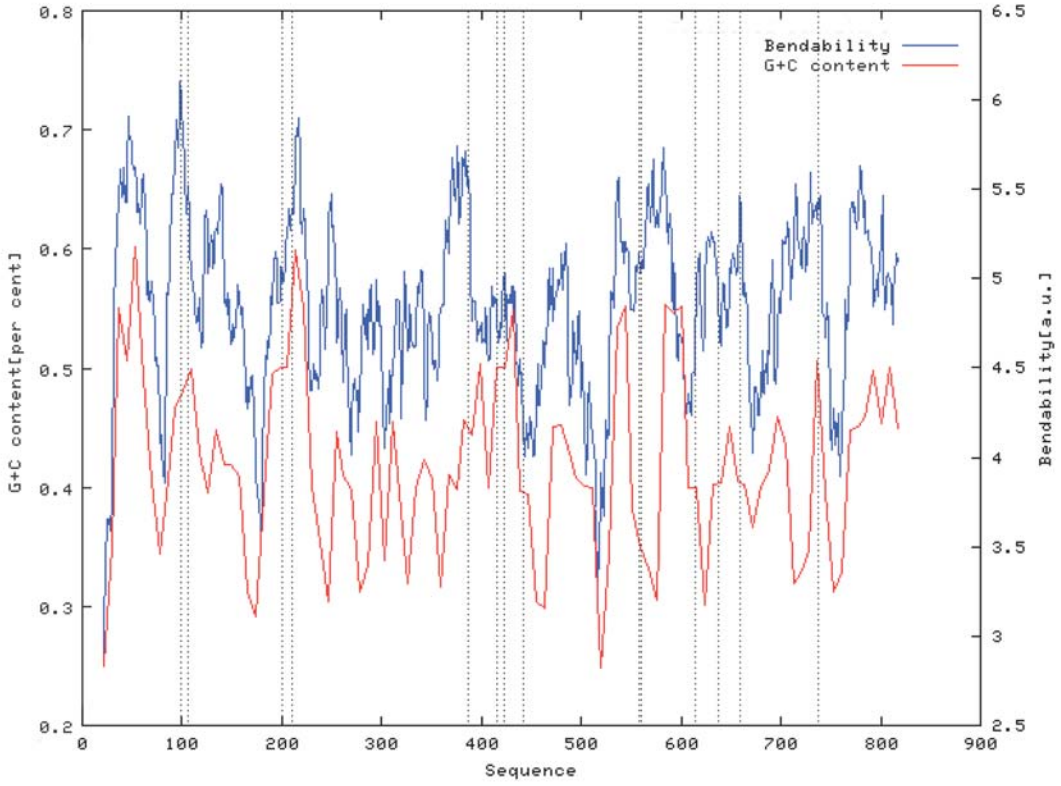

Bendability
G+C content
G+C content[per cent]
Bendability[a.u.]
Sequence
0.2
0.3
0.4
0.5
0.6
0.7
0.8
2.5
3
3.5
4
4.5
5
5.5
6
6.5
0
100
200
300
400
500
600

Figure 2.14 – Bendability and GC content plots for several Rd KW20 CDS with multiple recovered Tn5 insertions. **A.** HI0089 (40.6% G+C); **B.** HI0579 (40.7% G+C); **C.** HI0928 (42.7% G+C); **D.** HI0946.1 (42.1% G+C); **E.** HI1516-1518 (46.7% G+C); **F.** HI1520 (48.3% G+C). A 21-bp window was used to estimate bendability and GC content at each nucleotide surrounding the Tn5 insertion sites (blue and red lines respectively). The dashed vertical lines represent the center of the 9-bp repeat generated by the Tn5 insertion.

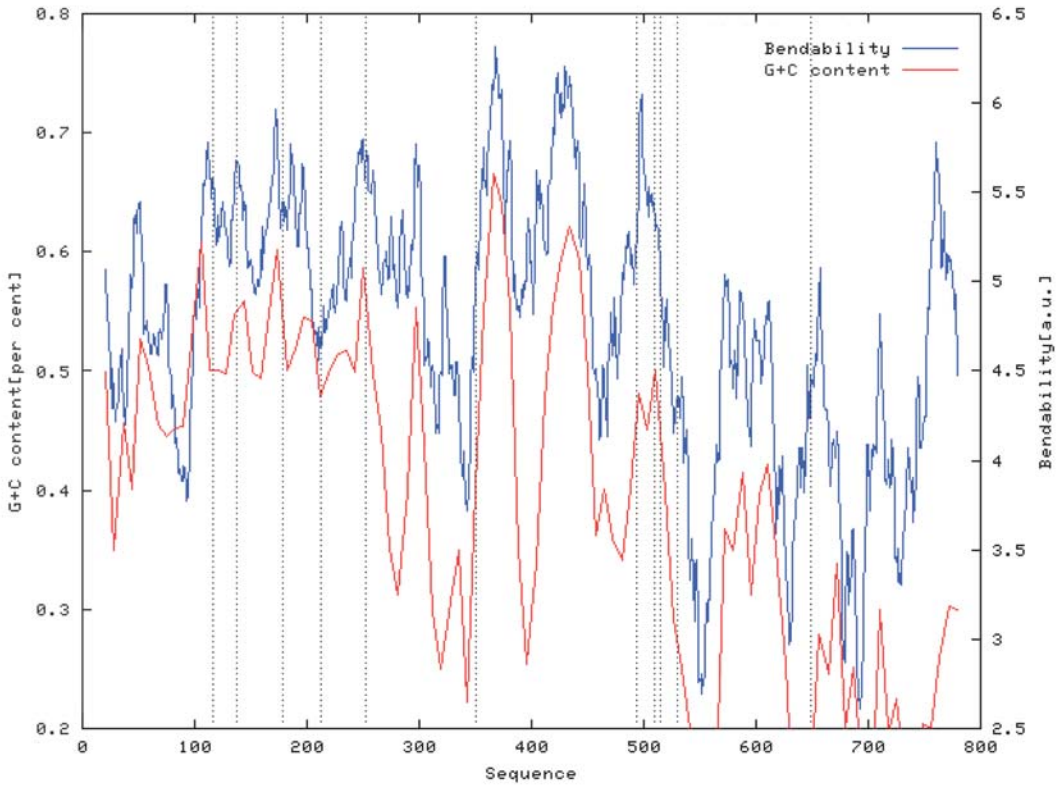
A


Bendability
G+C content
G+C content[per cent]
Bendability[a.u.]
Sequence
0.2
0.3
0.4
0.5
0.6
0.7
0.8
2.5
3
3.5
4
4.5
5
5.5
6
6.5
7
0
200
400
600
800
1000
1200
1400

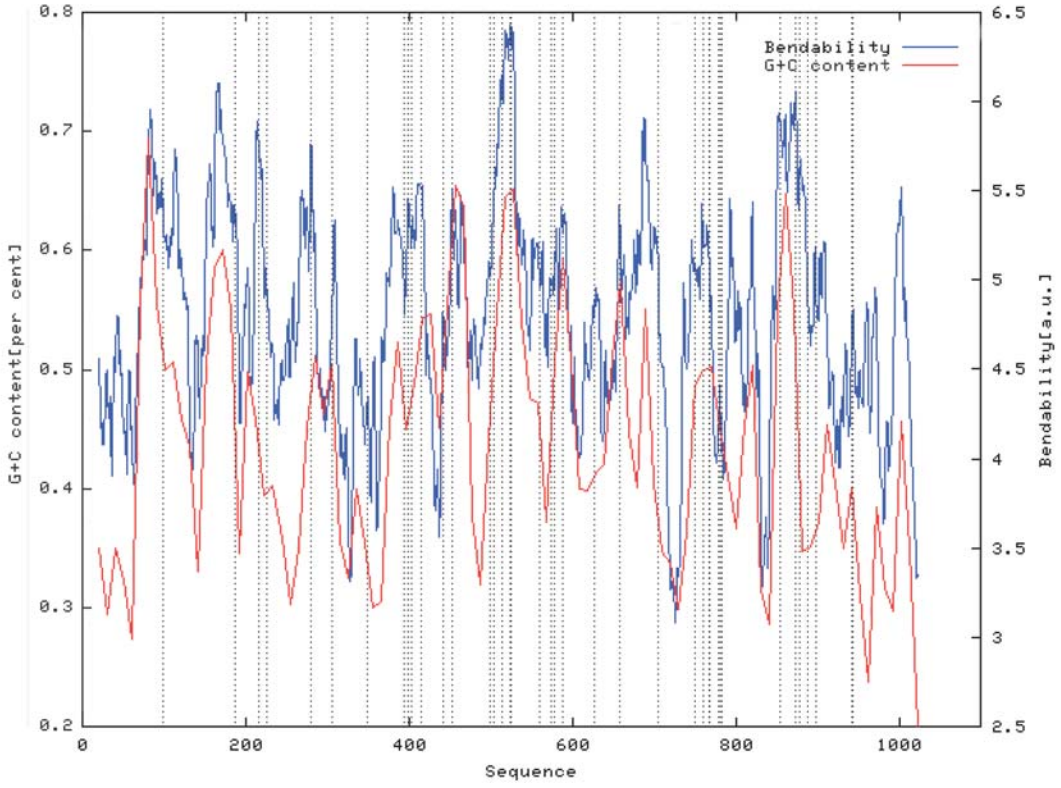
B


Bendability
G+C content
G+C content[per cent]
Bendability[a.u.]
Sequence
0.2
0.3
0.4
0.5
0.6
0.7
0.8
2.5
3
3.5
4
4.5
5
5.5
6
6.5
0
100
200
300
400
500
600
700
800
900

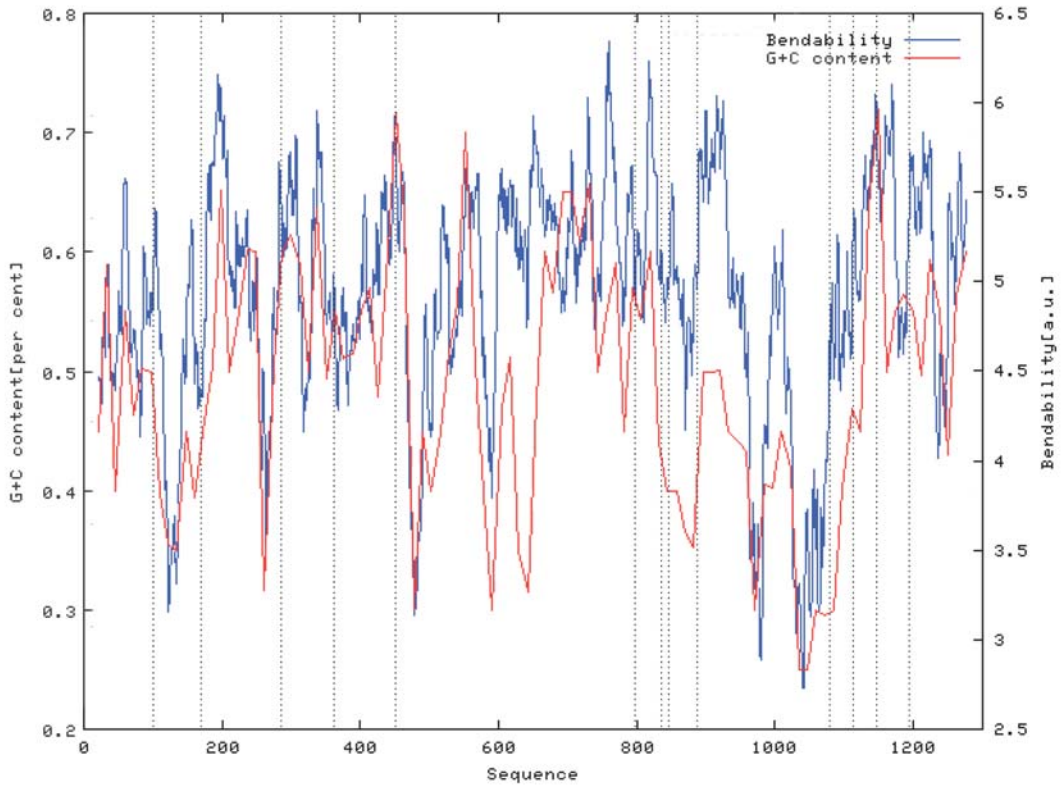
C


Bendability
G+C content
G+C content[per cent]
Bendability[a.u.]
Sequence
0.2
0.3
0.4
0.5
0.6
0.7
0.8
2.5
3
3.5
4
4.5
5
5.5
6
6.5
0
100
200
300
400
500
600
700
800

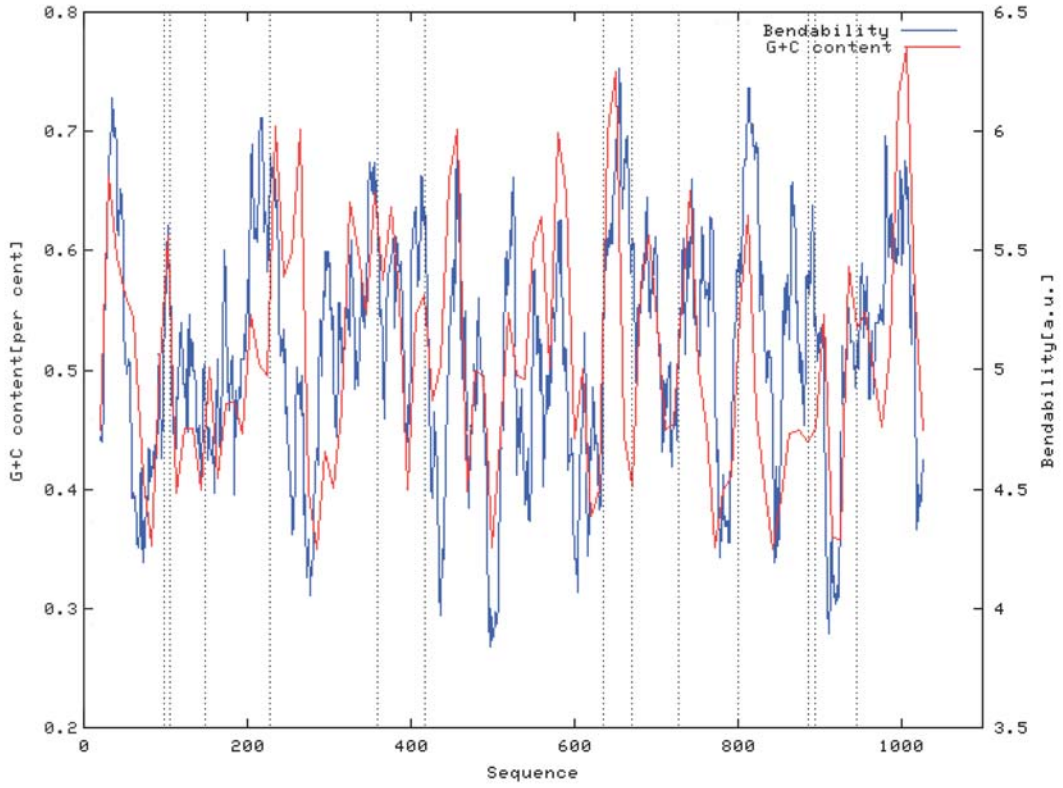
D


Bendability
G+C content
G+C content[per cent]
Bendability[a.u.]
Sequence
0.2
0.3
0.4
0.5
0.6
0.7
0.8
2.5
3
3.5
4
4.5
5
5.5
6
6.5
0
200
400
600
800
1000

E


Bendability
G+C content
G+C content[per cent]
Bendability[a.u.]
Sequence
0.2
0.3
0.4
0.5
0.6
0.7
0.8
2.5
3
3.5
4
4.5
5
5.5
6
6.5
0
200
400
600
800
1000
1200

F


Bendability
G+C content
G+C content[per cent]
Bendability[a.u.]
Sequence
0.2
0.3
0.4
0.5
0.6
0.7
0.8
3.5
4
4.5
5
5.5
6
6.5
0
200
400
600
800
1000

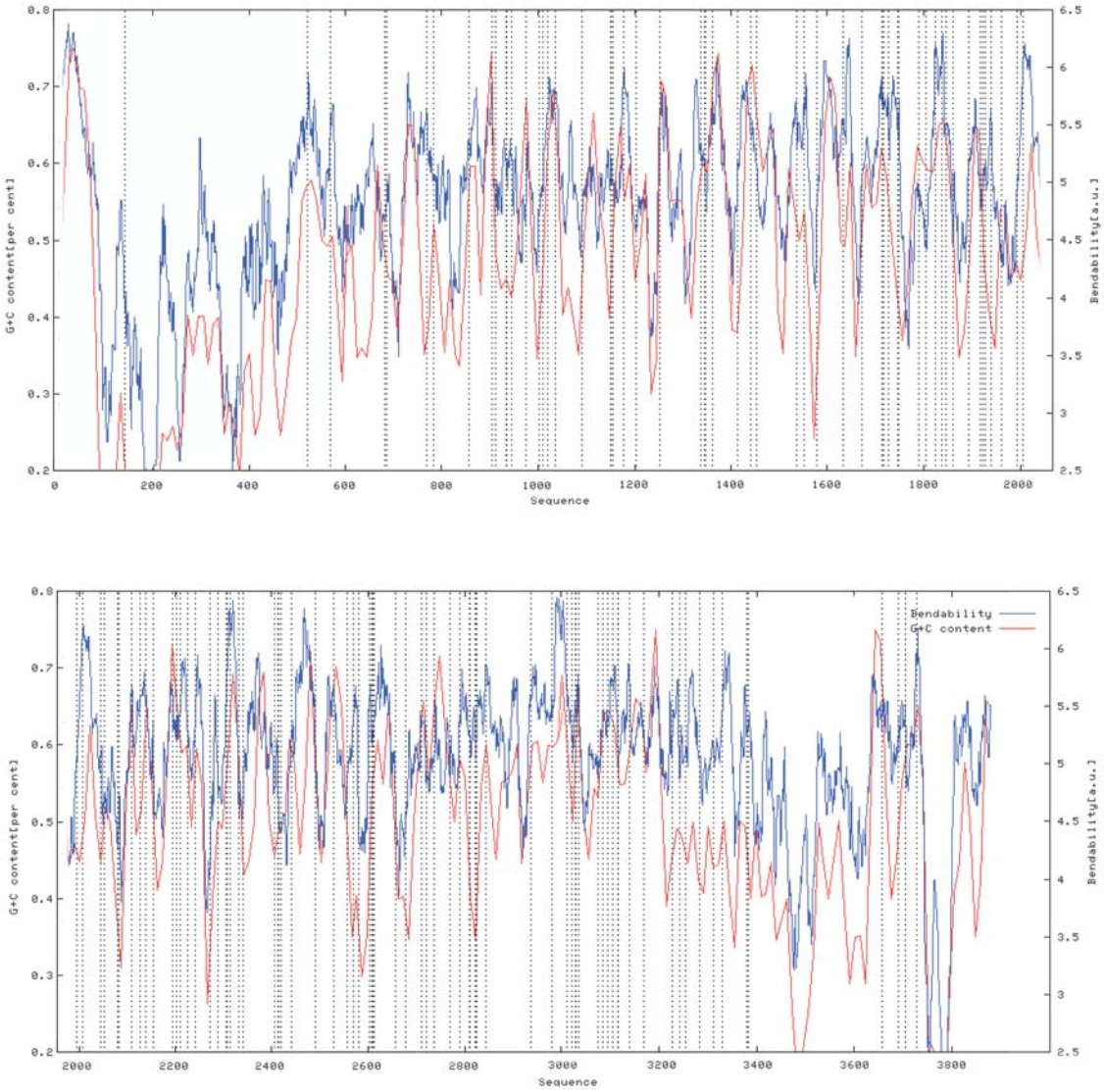


Figure 2.15 – Bendability and GC content plot of an Rd KW20 rRNA operon.

A 21-bp window was used to estimate bendability and GC content at each nucleotide surrounding the Tn5 insertion sites (blue and red lines respectively). The dashed vertical lines represent the center of the 9-bp repeat generated by the Tn5 insertion.

DISCUSSION

In this chapter, the results of a large-scale attempt at TIM in *H. influenzae* Rd KW20 are described. The goal of this research was to develop a system that would allow for the eventual creation of a comprehensive bank of transposon insertions into each of the annotated *H. influenzae* CDSs. These insertions would then be transformed into Rd KW20 for phenotypic analysis of the mutants to functionally characterize the non-essential genes. This system would be applicable to functional analysis of any naturally transformable organism. To facilitate this goal, a series of specialized cloning vectors were developed to reduce non-essential sequence in the plasmid backbone for lowered recovery of undesired transposon insertions. The pASC vectors utilized in this work were high-copy number plasmids derived from pUC19. Alpha-complementation by a β-galactosidase fragment encoded by *lacZα* on the plasmid is a valuable resource for discernment of colonies transformed with plasmids carrying an insert and those with the intact plasmid. While essential in this step of the process, the sequence necessary for production of β-galactosidase is unessential following the cloning procedure and provides a potential sink for transposon insertions. The vectors pASC13, pASC15 and pASC18 created in this work were designed to allow removal of *lacZα* sequence flanking of the chromosomal insert by restriction digestion with *Not*I and *Apa*I. Following post-cloning minimalization of pASC18, the nonessential regions are reduced to 11.6% of the total vector sequence. To further attempt to reduce the recovery of Tn5 insertions in the vector, adapter-mediated cloning into a minimized version of pASC18 lacking the *lacZα* sequence was employed. Additional mutagenesis of the vector reduced the nonessential

sequence to 8.3% of the total plasmid sequence, although this version of the vector was not employed for cloning and mutagenesis due to the end of this phase of the project.

In vitro transposition with Tn7 and Tn5 was employed with varied success in creating the library of mutagenic constructs. In total, 1290 plasmids were sequenced and the sites of 1426 Tn7 or Tn5 insertion events were mapped. Of these, 1024 (71.8%) were mapped to *H. influenzae* chromosomal DNA and were located in 246 of the 1710 Rd KW20 CDSs (14.4%).

While the initial *in vitro* transposition system used in this work, the GPS-1 Tn7-based system, resulted in the lowest frequency of recovered vector insertions, the repeated recovery of plasmids bearing an insertion in the Rd KW20 CDS HI0936 led to the eventual abandonment of this system. In total, 145 Tn7-mutated plasmids were recovered and 157 insertion sites were mapped. Of these sites, 134 (85.4%) were mapped in chromosomal DNA. These sites were located in 33 of the 1710 Rd KW20 CDSs and at 9 intergenic sites. Vector insertions accounted for only 14.7% of the insertions. Unfortunately, the small sample size precludes any analysis of the overall randomness of insertion into *H. influenzae* Rd KW20. Following initial success with the system, further attempts to mutate chromosomal libraries resulted in the return of multiple plasmids containing transposon insertions in the same site in HI0936. One of the two libraries in which this insertion was prevalent was concurrently mutated with Tn5 and transposon insertions were recovered in 42 different genes or intergenic regions (97 total chromosomal insertions). Since a lack of library complexity is not apparent, either human error, *i.e.* contaminated solutions, or a hot-spot for Tn7 exists in Rd KW20. The success

of the preliminary Tn5 mutagenesis led to the abandonment of the Tn7 system *in lieu* of attempting to solve this problem.

Insertions mapped with the *in vitro* Tn5 system constituted the bulk of the results of this study. In total, 1269 Tn5 insertion sites were mapped in 1145 plasmids. Of these, 890 were mapped in chromosomal DNA (70.1%) at 715 different sites in the Rd KW20 genome. These sites were located in 217 of the 1710 annotated CDSs (12.7%) and 28 intergenic sites. Additionally, 197 insertions were mapped to the rRNA operons, representing 22.1% of the chromosomal insertions. Vector insertions accounted for 29.9% of mapped insertions but the frequency of recovered insertions varied depending on which vector was used or library construction methodology. Many of these insertions were located in regions of the vector that could be characterized as essential since each mutagenic construct would require a working copy of these regions. This implies that vector concatemers, likely formed during either during the minimalization steps or adapter-mediated cloning into pASC18MIN, were frequently recovered.

The most successful strategy for reducing the frequency of vector Tn5 insertions was termed post-mutagenesis minimalization. This involved cloning *H. influenzae* genomic DNA into pASC18 followed immediately by *in vitro* mutagenesis. The non-essential regions of the vector were then removed by sequential *Apa*I and *Not*I digests. The most likely reason for the success of this method is that the majority of vectors harboring transposon insertions can be removed from the pool during the minimalization steps. While the vectors were subjected to CIP treatment prior to cloning, vector concatemers are likely to be present in the pool of plasmids subjected to mutagenesis. Additionally, plasmids that contain two or more vector backbones can occur in the pool

following minimalization. Since plasmids with chromosomal inserts less than 2-kbp were removed, a band corresponding to Tn5 insertions into vectors lacking a chromosomal insert is readily apparent and these can be removed from the pool during the minimalization procedure. This further reduces the recovery of Tn5 vector insertions to plasmids harboring a chromosomal insert. Therefore, recovery of vector insertions is limited to those in non-essential regions of vectors remaining after plasmid minimalization, in those in which a concatemer formed post-minimalization, or those in which the *Apa*I or *Not*I fragments of the vector were not properly excised.

The use of pASC18MIN was not successful in minimizing the number of vector insertions recovered. Adaptors were utilized to clone into the *Asc*I site of the vector that would remove the *Asc*I site if the proper insert was generated. A cycled ligation protocol that included *Asc*I should have reduced vector concatemerization since this event regenerates the original restriction site. The use of adapter-mediated cloning into pASC18MIN resulted in a dramatic increase (3.6-fold) in mapped vector insertions over the post-mutagenesis minimalization method. In theory, cloning directly into pASC18MIN (or pASC30) should result in the lowest possible yield of vector Tn5 insertions. Even if the assumption is made that the cycled ligation step used was completely efficient and followed by an additional *Asc*I digestion, plasmids containing two or more vector backbones can still arise by two means; *Asc*I-resistant vector concatemers can form from damaged or blunted overhangs and complex plasmids with two inserts and two vector backbones. Simple vector concatemers should have been easily removed by gel electrophoresis as they would be smaller than a vector containing a chromosomal insert but removal of the more complex plasmids would be difficult. In

retrospect, the rate of pASC18::Tn5 recovery obtained with post-mutagenesis minimalization was sufficiently low enough to justify the use of this method in any future TIM work. Alternatively, the Gateway cloning technology may be a better methodology (106); however, it was not available at the time this work was performed. The Gateway system has been successfully used with Tn5 mutagenesis in order to reduce recovery of vector insertions (90). In this system, the mutagenesis step is followed by recombination of the chromosomal insert from the original cloning vector into a new vector, but prevents incorporation of the original vector sequence into the new plasmid. Therefore, Tn5 vector insertions can be completely eliminated from the pools of mutants to be mapped.

Previous studies on Tn5 insertion preferences have defined the consensus 9-bp sequence duplicated by Tn5 insertion as 5'-GYYYWRRRC (39, 90). While this consensus sequence is based on the analysis of many thousands of insertions, individual sites in the Shevchenko study averaged only 61% accordance with this sequence. The results presented in this study are comparable to the Shevchenko study and other previous studies. Only six insertion sites from this study matched the consensus sequence. The data from this study implicate a possible GC-content bias for transposition of Tn5. One finding demonstrates a statistically significant increase in the average GC content of the 40-bp surrounding the chromosomal Tn5 insertion sites from that of Rd KW20. In contrast, the average GC content of the Tn5 insertion sites in vector DNA sequences was not significantly higher than that of the vector as a whole. The %G+C increase for chromosomal insertions is probably the result a related bias encountered in this work. Tn5 insertions were more likely to be found in CDSs with a higher than average GC

content and in the rRNA operons, which are close to 50% G+C. Two possibilities can account for this bias: the results may simply reflect that the chromosomal libraries were biased towards higher GC DNA fragments or alternatively, the mutagenesis was biased by selectivity of the transposase for targets in more GC-rich areas. While it is impossible to prove conclusively that the libraries themselves were not biased, examination of the restriction digest fragments used for library construction indicated that any bias was unlikely the result of the choices of restriction enzymes. Additionally, analysis of the predicted restriction fragments on which Tn5 insertions were mapped indicates that many CDSs with low GC content were likely present in the libraries but were not targets of Tn5 insertions. It would appear that regions of balanced AT/GC ratios are preferred for Tn5 insertions. The extreme difference in the average GC contents of *H. influenzae* chromosomal DNA and the pASC vectors (38% and 52% respectively) may also explain the higher-than-expected numbers of vector insertions.

The results of the Shevchenko study indicated that there was a bias in Tn5 insertions but that it could not be explained by GC content (90). There are reasons why a GC bias might be easier to recognize from this work. From the graphs presented in their study, it appears that the average GC content of the cDNA clones was approximately 55% and approximately 1.9-kp in length. Each clone was also individually mutated. If the GC content of each clone was fairly homogenous over the length of the fragment and there were no competing clones within the same mutagenesis pool, any bias against insertion into AT-rich clones or regions would not be apparent.

The lack of a clear consensus sequence argues that recognition of target sites by the transposase is through other factors such as structure, availability or topology.

Previous studies of target sites in the P element and *mariner*-class elements indicated that properties of the DNA such as A-philicity, bendability, protein-induced deformability or B-DNA twist may be involved in target site selection (59, 104). Plots for several of these properties indicated a qualitative difference between the signals generated at the Tn5 insertion sites recovered in Rd KW20 chromosomal DNA and that of randomly generated DNA sequences from this organism. This was also true for insertions into the vector sequence, although the difference between these insertions and randomly generated vector sequences did not appear as large. Interestingly, the characteristics at the Tn5 insertion sites are very similar to the *D. melanogaster* P element sites and suggest a common origin. While the target duplication is only 8-bp in P element transposition, it has a consensus sequence (5'-GTYYRRAC) very similar to that of Tn5. P-element transposition is more common in regions of higher GC content and the profiles for bendability, A-philicity, PID and B-DNA twist are very similar to that observed with Tn5 (59).

Of the properties examined, the ability of the DNA to bend at the insertion site may be the most biologically significant. Increases in the predicted bendability of the DNA surrounding the insertion sites were prominent for the P element and Tc1/*mariner* elements (59, 104). Recent studies of Tn10 insertions have shown that DNA is bent by the binding of the transposase and that bendability is an important determinant in target site selection (76). Additionally, DNA bending by the transposase or accessory proteins have been demonstrated for Sleeping Beauty, IS231A, Tn7 and Mu phage transposition (8, 22, 41, 111). While individual trinucleotide GC values correlated poorly with bendability, overall bendability of the DNA is highly correlated with GC content (105).

The skewed distribution towards DNAs of higher GC content encountered in this study is understandable if bendability of DNA is an important factor in target selection by the Tn5 transposase.

Overall, these results suggest that continued use of Tn5 mutagenesis of *H. influenzae* DNA is unlikely to be successful in saturating the genome if the current methodology for library creation is continued. The use of pooled libraries of mixed GC content likely requires the mapping of extremely large numbers of transposon insertions in order to obtain mutations in every CDS. Additionally, the higher GC content of the minimized vector might prove too inviting a target to mutate the most AT-rich chromosomal fragments and Gateway cloning technology might not be able to alleviate the low frequency of insertions into these inserts. This problem likely requires the use of a different transposon for mutagenesis of AT-rich chromosomal sequences.

Since the inception of this work, other studies have been published that used large-scale functional genomics techniques to study *H. influenzae*. Several of these studies utilized transposon mutagenesis. The largest of these involved GAMBIT analysis to determine essential genes in *H. influenzae* (4). This study identified 670 putative essential genes and 538 non-essential genes. In essence, the methodology involves *in vitro* mutagenesis of defined PCR products followed by transformation and selection for recombinant mutants. Genetic footprinting is then used to identify genes in which transposon insertions are not recovered. Unfortunately, this technique has serious flaws that result in mischaracterization of genes as essential. An assumption is made that the transposon insertions are random and evenly distributed. The authors putatively assign essentiality to genes in which mutations are not recovered without prior analysis of

whether insertions in that gene were present in the mutagenesis pool. As a result, a large number of genes deemed essential by this technique have been previously shown to be nonessential, including most of the genes involved in competence and transformation in *H. influenzae.* This flaw could be easily remedied by footprinting the mutagenesis pools for comparison to the data following mutant recovery. Transposon-based signature-tagged mutagenesis has also been employed to study *H. influenzae* (49). That study was able to identify 25 genes required for *H. influenzae* invasive disease. Most of the *H. influenzae* proteomics studies have involved preliminary analysis of the proteome (53, 56, 101). One study utilized 2D-electrophoresis to study changes in protein expression patterns related to exposure to various antibiotics (25). A second study combined microarrays and proteomics to analyze changes in expression resulting from exposure to novobiocin and ciprofloxacin (36). This latter study remains the only published use of microarray technology to study *H. influenzae* gene expression.

In the current study, the use of *in vitro* mutagenesis is described, primarily using Tn5, to create a comprehensive library of transposon-disrupted genes for functional analysis of the *H. influenzae* genome. While this work was successful in disrupting 246 of the 1710 annotated CDSs, the number of sequencing reactions required to achieve this result was extraordinarily high. This would appear to be due to a previously uncharacterized bias against Tn5 transposition into AT-rich DNA. Future work using TIM in *H. influenzae* can benefit from the insights generated from this study. First, mutagenesis of large, relatively undefined pools of DNA is an inefficient method in the absence of a transposon system lacking any sequence, structural or topological biases. Since such a system may not exist, mutagenesis of clearly defined pools of chromosomal

DNAs would be a much more efficient method since it allows easy removal of fragments where saturation by the transposon had been achieved. Secondly, success in saturating the complete genome might require the use of multiple transposons with varied biases. Nevertheless, the protocols, software and vectors developed for this work would be useful for functional genomic studies of organisms with a higher GC genomic content. In addition, the minimal vectors developed for this work would be highly useful for directed deletion mutagenesis using an inverse PCR strategy.

While *H. influenzae* Rd KW20 was the first free-living organism to be completely sequenced, functional genomic analysis has lagged behind that of other model organisms. This is likely due in part to the reduced incidence of fatalities caused by *H. influenzae* meningitis, but NTHi continues to be a significant human pathogen. For this reason, it is important to functionally characterize the *H. influenzae* genome. The number of insertions recovered in this study was sufficient to justify an attempt to fulfill the second goal of this project, *i.e.* transformation of the insertions into *H. influenzae* Rd KW20 with subsequent use of the mutant bank to screen for phenotypic defects related to the disruption of the genes. To this end, the following chapters describe the results of utilizing the transposon libraries generated in this work to search for novel genes mediating transformation in *H. influenzae*. In addition, *in silico* and transcriptional analyses, combined with directed mutagenesis, were utilized to further characterize the competence regulon and transformation machinery in this organism.

REFERENCES (PART I)

1. **Adams, W. G., K. A. Deaver, S. L. Cochi, B. D. Plikaytis, E. R. Zell, C. V. Broome, and J. D. Wenger**. 1993. Decline of childhood *Haemophilus influenzae* type b (Hib) disease in the Hib vaccine era. JAMA **269**:221-226.

2. **Aebersold, R. and M. Mann**. 2003. Mass spectrometry-based proteomics. Nature **422**:198-207.

3. **Akerley, B. J., E. J. Rubin, A. Camilli, D. J. Lampe, H. M. Robertson, and J. J. Mekalanos**. 1998. Systematic identification of essential genes by *in vitro* mariner mutagenesis. Proc. Natl. Acad. Sci. U. S. A. **95**:8927-32.

4. **Akerley, B. J., E. J. Rubin, V. L. Novick, K. Amaya, N. Judson, and J. J. Mekalanos**. 2002. A genome-scale analysis for identification of genes required for growth or survival of *Haemophilus influenzae*. Proc. Natl. Acad. Sci. U. S. A. **99**:966-971.

5. **Albano, M., J. Hahn, and D. Dubnau**. 1987. Expression of competence genes in *Bacillus subtilis*. J. Bacteriol. **169**:3110-3117.

6. **Alexander, H. and G. Leidy**. 1951. Determination of inherited traits of *Haemophilus influenzae* by desoxyribonucleic acid fractions isolated from type-specific cells. J. Exp. Med. **93**:345-359.

7. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman**. 1990. Basic local alignment search tool. J. Mol. Biol. **215**:403-410.

8. **Arciszewska, L. K. and N. L. Craig**. 1991. Interaction of the Tn7-encoded transposition protein TnsB with the ends of the transposon. Nucleic Acids Res. **19**:5021-5029.

9. **Ariazi, E. A. and M. N. Gould**. 1996. Consecutive cycles of precise, unidirectional 14-bp deletions using a *Bse*RI/*Bsg*I trimming plasmid. Biotechniques **20**:446-1.

10. **Aubrey, R. and C. Tang**. 2003. The pathogenesis of disease due to type b *Haemophilus influenzae*. Methods Mol. Med. **71**:29-50.

11. **Ausubel, F. M.** 1987. Current protocols in molecular biology. Greene Publishing Associates, Brooklyn, N. Y.

12. **Bainton, R. J., K. M. Kubo, J. N. Feng, and N. L. Craig**. 1993. Tn7 transposition: target DNA recognition is mediated by multiple Tn7-encoded proteins in a purified *in vitro* system. Cell **72**:931-943.

13. **Bairoch, A.** 1991. PROSITE: a dictionary of sites and patterns in proteins. Nucleic Acids Res. **19 Suppl**:2241-2245.

14. **Behr, M. A., M. A. Wilson, W. P. Gill, H. Salamon, G. K. Schoolnik, S. Rane, and P. M. Small**. 1999. Comparative genomics of BCG vaccines by whole-genome DNA microarray. Science **284**:1520-1523.

15. **Berka, R. M., J. Hahn, M. Albano, I. Draskovic, M. Persuh, X. Cui, A. Sloma, W. Widner, and D. Dubnau**. 2002. Microarray analysis of the *Bacillus subtilis* K-state: genome-wide expression changes dependent on ComK. Mol. Microbiol. **43**:1331-1345.

16. **Betts, J. C.** 2002. Transcriptomics and proteomics: tools for the identification of novel drug targets and vaccine candidates for tuberculosis. IUBMB Life **53**:239-242.

17. **Bidlingmaier, S. and M. Snyder**. 2002. Large-scale identification of genes important for apical growth in *Saccharomyces cerevisiae* by directed allele replacement technology (DART) screening. Funct. Integr. Genomics **1**:345-356.

18. **Brown, P. O. and D. Botstein**. 1999. Exploring the new world of the genome with DNA microarrays. Nat. Genet. **21**:33-7.

19. **Brukner, I., R. Sanchez, D. Suck, and S. Pongor**. 1995. Trinucleotide models for DNA bending propensity: comparison of models based on DNaseI digestion and nucleosome packaging data. J. Biomol. Struct. Dyn. **13**:309-317.

20. **Campoy, S., M. Fontes, S. Padmanabhan, P. Cortes, M. Llagostera, and J. Barbe**. 2003. LexA-independent DNA damage-mediated induction of gene expression in *Myxococcus xanthus*. Mol. Microbiol. **49**:769-781.

21. **Casadaban, M. J.** 1976. Transposition and fusion of the *lac* genes to selected promoters in *Escherichia coli* using bacteriophage lambda and *Mu*. J. Mol. Biol. **104**:541-555.

22. **Chaconas, G.** 1999. Studies on a "jumping gene machine": higher-order nucleoprotein complexes in *Mu* DNA transposition. Biochem. Cell Biol. **77**:487-491.

23. **Chia, J. S., Y. Y. Lee, P. T. Huang, and J. Y. Chen**. 2001. Identification of stress-responsive genes in *Streptococcus mutans* by differential display reverse transcription-PCR. Infect. Immun. **69**:2493-2501.

24. **DeRisi, J. L., V. R. Iyer, and P. O. Brown**. 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. Science **278**:680-686.

25. **Evers, S., K. Di Padova, M. Meyer, H. Langen, M. Fountoulakis, W. Keck, and C. P. Gray**. 2001. Mechanism-related changes in the gene transcription and

protein synthesis patterns of *Haemophilus influenzae* after treatment with transcriptional and translational inhibitors. Proteomics **1**:522-544.

26. **Ferguson, P. L. and R. D. Smith**. 2003. Proteome analysis by mass spectrometry. Annu. Rev. Biophys Biomol. Struct. **32**:399-424.

27. **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.

28. **Foxwell, A. R., J. M. Kyd, and A. W. Cripps**. 1998. Nontypeable *Haemophilus influenzae*: pathogenesis and prevention. Microbiol. Mol. Biol. Rev. **62**:294-308.

29. **Fuller, T. E., M. J. Kennedy, and D. E. Lowery**. 2000. Identification of *Pasteurella multocida* virulence genes in a septicemic mouse model using signature-tagged mutagenesis. Microb. Pathog. **29**:25-38.

30. **Fuller, T. E., S. Martin, J. F. Teel, G. R. Alaniz, M. J. Kennedy, and D. E. Lowery**. 2000. Identification of *Actinobacillus pleuropneumoniae* virulence genes using signature-tagged mutagenesis in a swine infection model. Microb. Pathog. **29**:39-51.

31. **Gabrielian, A., A. Simoncsits, and S. Pongor**. 1996. Distribution of bending propensity in DNA sequences. FEBS Lett. **393**:124-130.

32. **Garraway, L. A., L. R. Tosi, Y. Wang, J. B. Moore, D. E. Dobson, and S. M. Beverley**. 1997. Insertional mutagenesis by a modified *in vitro* Ty1 transposition system. Gene **198**:27-35.

33. **Gerdes, S. Y., M. D. Scholle, M. D'Souza, A. Bernal, M. V. Baev, M. Farrell, O. V. Kurnasov, M. D. Daugherty, F. Mseeh, B. M. Polanuyer, J. W. Campbell, S. Anantha, K. Y. Shatalin, S. A. Chowdhury, M. Y. Fonstein, and A. L. Osterman**. 2002. From genetic footprinting to antimicrobial drug targets: examples in cofactor biosynthetic pathways. J. Bacteriol. **184**:4555-4572.

34. **Giaever, G., A. M. Chu, L. Ni, C. Connelly, L. Riles, S. Veronneau, S. Dow, A. Lucau-Danila, K. Anderson, B. Andre, A. P. Arkin, A. Astromoff, M. El Bakkoury, R. Bangham, R. Benito, S. Brachat, S. Campanaro, M. Curtiss, K. Davis, A. Deutschbauer, K. D. Entian, P. Flaherty, F. Foury, D. J. Garfinkel, M. Gerstein, D. Gotte, U. Guldener, J. H. Hegemann, S. Hempel, Z. Herman, D. F. Jaramillo, D. E. Kelly, S. L. Kelly, P. Kotter, D. LaBonte, D. C. Lamb, N. Lan, H. Liang, H. Liao, L. Liu, C. Luo, M. Lussier, R. Mao, P. Menard, S. L. Ooi, J. L. Revuelta, C. J. Roberts, M. Rose, P. Ross-Macdonald, B. Scherens, G. Schimmack, B. Shafer, D. D. Shoemaker, S. Sookhai-Mahadeo, R. K. Storms, J. N. Strathern, G. Valle, M. Voet, G. Volckaert, C. Y. Wang, T. R. Ward, J. Wilhelmy, E. A. Winzeler, Y. Yang, G. Yen, E. Youngman, K. Yu, H. Bussey, J. D. Boeke, M. Snyder, P. Philippsen, R. W. Davis, and M.**

Johnston. 2002. Functional profiling of the *Saccharomyces cerevisiae* genome. Nature **418**:387-391.

35. **Ginzinger, D. G.** 2002. Gene quantification using real-time quantitative PCR: an emerging technology hits the mainstream. Exp. Hematol. **30**:503-512.

36. **Gmuender, H., K. Kuratli, K. Di Padova, C. P. Gray, W. Keck, and S. Evers**. 2001. Gene expression changes triggered by exposure of *Haemophilus influenzae* to novobiocin or ciprofloxacin: combined transcription and translation analysis. Genome Res. **11**:28-42.

37. **Goryshin, I. Y., J. Jendrisak, L. M. Hoffman, R. Meis, and W. S. Reznikoff**. 2000. Insertional transposon mutagenesis by electroporation of released Tn5 transposition complexes. Nat. Biotechnol. **18**:97-100.

38. **Goryshin, I. Y., J. A. Miller, Y. V. Kil, V. A. Lanzov, and W. S. Reznikoff**. 1998. Tn5/IS50 target recognition. Proc. Natl. Acad. Sci. U. S. A. **95**:10716-10721.

39. **Goryshin, I. Y. and W. S. Reznikoff**. 1998. Tn5 *in vitro* transposition. J. Biol. Chem. **273**:7367-7374.

40. **Greenbaum, D., C. Colangelo, K. Williams, and M. Gerstein**. 2003. Comparing protein abundance and mRNA expression levels on a genomic scale. Genome Biol. **4**:117.

41. **Hallet, B., R. Rezsohazy, J. Mahillon, and J. Delcour**. 1994. IS231A insertion specificity: consensus sequence and DNA bending at the target site. Mol. Microbiol. **14**:131-139.

42. **Hamer, L., K. Adachi, M. V. Montenegro-Chamorro, M. M. Tanzer, S. K. Mahanty, C. Lo, R. W. Tarpey, A. R. Skalchunes, R. W. Heiniger, S. A. Frank, B. A. Darveaux, D. J. Lampe, T. M. Slater, L. Ramamurthy, T. M. DeZwaan, G. H. Nelson, J. R. Shuster, J. Woessner, and J. E. Hamer**. 2001. Gene discovery and gene function assignment in filamentous fungi. Proc. Natl. Acad. Sci. U. S. A. **98**:5110-5115.

43. **Hanna, M. N., R. J. Ferguson, Y. H. Li, and D. G. Cvitkovitch**. 2001. *uvrA* is an acid-inducible gene involved in the adaptive response to low pH in *Streptococcus mutans*. J. Bacteriol. **183**:5964-5973.

44. **Hare, R. S., S. S. Walker, T. E. Dorman, J. R. Greene, L. M. Guzman, T. J. Kenney, M. C. Sulavik, K. Baradaran, C. Houseweart, H. Yu, Z. Foldes, A. Motzer, M. Walbridge, G. H. Shimer, Jr., and K. J. Shaw**. 2001. Genetic footprinting in bacteria. J. Bacteriol. **183**:1694-1706.

45. **Harrington, C. A., C. Rosenow, and J. Retief**. 2000. Monitoring gene expression using DNA microarrays. Curr. Opin. Microbiol. **3**:285-291.

46. **Hayes, F.** 2003. Transposon-based strategies for microbial functional genomics and proteomics. Annu. Rev. Genet. **37**:3-29.

47. **Helmann, J. D., M. F. Wu, P. A. Kobel, F. J. Gamo, M. Wilson, M. M. Morshedi, M. Navre, and C. Paddon**. 2001. Global transcriptional response of *Bacillus subtilis* to heat shock. J. Bacteriol. **183**:7318-7328.

48. **Hensel, M., J. E. Shea, C. Gleeson, M. D. Jones, E. Dalton, and D. W. Holden**. 1995. Simultaneous identification of bacterial virulence genes by negative selection. Science **269**:400-3.

49. **Herbert, M. A., S. Hayes, M. E. Deadman, C. M. Tang, D. W. Hood, and E. R. Moxon**. 2002. Signature Tagged Mutagenesis of Haemophilus influenzae identifies genes required for in vivo survival. Microb. Pathog. **33**:211-223.

50. **Jarmer, H., R. Berka, S. Knudsen, and H. H. Saxild**. 2002. Transcriptome analysis documents induced competence of *Bacillus subtilis* during nitrogen limiting conditions. FEMS Microbiol. Lett. **206**:197-200.

51. **Jonczyk, P. and A. Nowicka**. 1996. Specific in vivo protein-protein interactions between *Escherichia coli* SOS mutagenesis proteins. J. Bacteriol. **178**:2580-2585.

52. **Kobayashi, K., M. Ogura, H. Yamaguchi, K. Yoshida, N. Ogasawara, T. Tanaka, and Y. Fujita**. 2001. Comprehensive DNA microarray analysis of *Bacillus subtilis* two-component regulatory systems. J. Bacteriol. **183**:7365-7370.

53. **Kolker, E., S. Purvine, M. Y. Galperin, S. Stolyar, D. R. Goodlett, A. I. Nesvizhskii, A. Keller, T. Xie, J. K. Eng, E. Yi, L. Hood, A. F. Picone, T. Cherny, B. C. Tjaden, A. F. Siegel, T. J. Reilly, K. S. Makarova, B. O. Palsson, and A. L. Smith**. 2003. Initial proteome analysis of model microorganism *Haemophilus influenzae* strain Rd KW20. J. Bacteriol. **185**:4593-4602.

54. **Lamberg, A., S. Nieminen, M. Qiao, and H. Savilahti**. 2002. Efficient insertion mutagenesis strategy for bacterial genomes involving electroporation of *in vitro*-assembled DNA transposition complexes of bacteriophage mu. Appl. Environ. Microbiol. **68**:705-712.

55. **Lampe, D. J., M. E. Churchill, and H. M. Robertson**. 1996. A purified mariner transposase is sufficient to mediate transposition *in vitro*. EMBO J. **15**:5470-5479.

56. **Langen, H., B. Takacs, S. Evers, P. Berndt, H. W. Lahm, B. Wipf, C. Gray, and M. Fountoulakis**. 2000. Two-dimensional map of the proteome of *Haemophilus influenzae*. Electrophoresis **21**:411-429.

57. **Liang, P., L. Averboukh, and A. B. Pardee**. 1993. Distribution and cloning of eukaryotic mRNAs by means of differential display: refinements and optimization. Nucleic Acids Res. **21**:3269-3275.

58. **Liang, P. and A. B. Pardee**. 1992. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science **257**:967-71.

59. **Liao, G. C., E. J. Rehm, and G. M. Rubin**. 2000. Insertion site preferences of the P transposable element in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. U. S. A. **97**:3347-3351.

60. **Lockhart, D. J. and E. A. Winzeler**. 2000. Genomics, gene expression and DNA arrays. Nature **405**:827-836.

61. **Lund, A. H., M. Duch, and F. S. Pedersen**. 1996. Increased cloning efficiency by temperature-cycle ligation. Nucleic Acids Res. **24**:800-801.

62. **Maniatis, T., E. F. Fritsch, and J. Sambrook**. 1982. Molecular cloning a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

63. **Martin, K. J. and A. B. Pardee**. 1999. Principles of differential display. Methods Enzymol. **303**:234-58.

64. **Marykwas, D. L., S. A. Schmidt, and H. C. Berg**. 1996. Interacting components of the flagellar motor of *Escherichia coli* revealed by the two-hybrid system in yeast. J. Mol. Biol. **256**:564-576.

65. **Mason, K. M., R. S. Munson, Jr., and L. O. Bakaletz**. 2003. Nontypeable *Haemophilus influenzae* gene expression induced *in vivo* in a chinchilla model of otitis media. Infect. Immun. **71**:3454-3462.

66. **McHugh, J. P., F. Rodriguez-Quinones, H. Abdul-Tehrani, D. A. Svistunenko, R. K. Poole, C. E. Cooper, and S. C. Andrews**. 2003. Global iron-dependent gene regulation in *Escherichia coli*. A new mechanism for iron homeostasis. J. Biol. Chem. **278**:29478-29486.

67. **Mewes, H. W., K. Albermann, M. Bahr, D. Frishman, A. Gleissner, J. Hani, K. Heumann, K. Kleine, A. Maierl, S. G. Oliver, F. Pfeiffer, and A. Zollner**. 1997. Overview of the yeast genome. Nature **387**:7-65.

68. **Murzin, A. G., S. E. Brenner, T. Hubbard, and C. Chothia**. 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. **247**:536-540.

69. **Ogunniyi, A. D., P. Giammarinaro, and J. C. Paton**. 2002. The genes encoding virulence-associated proteins and the capsule of *Streptococcus pneumoniae* are upregulated and differentially expressed *in vivo*. Microbiology **148**:2045-2053.

70. **Ogura, M., H. Yamaguchi, K. Kobayashi, N. Ogasawara, Y. Fujita, and T. Tanaka**. 2002. Whole-genome analysis of genes regulated by the *Bacillus subtilis* competence transcription factor ComK. J. Bacteriol. **184**:2344-2351.

71. **Ogura, M., H. Yamaguchi, K. Yoshida, Y. Fujita, and T. Tanaka**. 2001. DNA microarray analysis of *Bacillus subtilis* DegU, ComA and PhoP regulons: an approach to comprehensive analysis of *B.subtilis* two-component regulatory systems. Nucleic Acids Res. **29**:3804-3813.

72. **Patterson, S. D. and R. H. Aebersold**. 2003. Proteomics: the first decade and beyond. Nat. Genet. **33 Suppl**:311-323.

73. **Paustian, M. L., B. J. May, and V. Kapur**. 2001. *Pasteurella multocida* gene expression in response to iron limitation. Infect. Immun. **69**:4109-4115.

74. **Paustian, M. L., B. J. May, and V. Kapur**. 2002. Transcriptional response of *Pasteurella multocida* to nutrient limitation. J. Bacteriol. **184**:3734-3739.

75. **Peterson, S., R. T. Cline, H. Tettelin, V. Sharov, and D. A. Morrison**. 2000. Gene expression analysis of the *Streptococcus pneumoniae* competence regulons by use of DNA microarrays. J. Bacteriol. **182**:6192-6202.

76. **Pribil, P. A. and D. B. Haniford**. 2003. Target DNA bending is an important specificity determinant in target site selection in Tn10 transposition. J. Mol. Biol. **330**:247-259.

77. **Rain, J. C., L. Selig, H. De Reuse, V. Battaglia, C. Reverdy, S. Simon, G. Lenzen, F. Petel, J. Wojcik, V. Schachter, Y. Chemama, A. Labigne, and P. Legrain**. 2001. The protein-protein interaction map of *Helicobacter pylori*. Nature **409**:211-215.

78. **Rajeevan, M. S., S. D. Vernon, N. Taysavang, and E. R. Unger**. 2001. Validation of array-based gene expression profiles by real-time (kinetic) RT-PCR. J. Mol. Diagn. **3**:26-31.

79. **Reznikoff, W. S. and K. P. Thornton**. 1972. Isolating tryptophan regulatory mutants in *Escherichia coli* by using a *trp-lac* fusion strain. J. Bacteriol. **109**:526-532.

80. **Richmond, C. S., J. D. Glasner, R. Mau, H. Jin, and F. R. Blattner**. 1999. Genome-wide expression profiling in *Escherichia coli* K-12. Nucleic Acids Res. **27**:3821-3835.

81. **Rimini, R., B. Jansson, G. Feger, T. C. Roberts, M. de Francesco, A. Gozzi, F. Faggioni, E. Domenici, D. M. Wallace, N. Frandsen, and A. Polissi**. 2000. Global analysis of transcription kinetics during competence development in *Streptococcus pneumoniae* using high density DNA arrays. Mol. Microbiol. **36**:1279-1292.

82. **Ross-Macdonald, P., P. S. Coelho, T. Roemer, S. Agarwal, A. Kumar, R. Jansen, K. H. Cheung, A. Sheehan, D. Symoniatis, L. Umansky, M. Heidtman, F. K. Nelson, H. Iwasaki, K. Hager, M. Gerstein, P. Miller, G. S.**

Roeder, and M. Snyder. 1999. Large-scale analysis of the yeast genome by transposon tagging and gene disruption. Nature **402**:413-418.

83. **Rudi, K., H. K. Nogva, K. Naterstad, S. M. Dromtorp, S. Bredholt, and A. Holck**. 2003. Subtyping *Listeria monocytogenes* through the combined analyses of genotype and expression of the *hlyA* virulence determinant. J. Appl. Microbiol. **94**:720-732.

84. **Saha, S., A. B. Sparks, C. Rago, V. Akmaev, C. J. Wang, B. Vogelstein, K. W. Kinzler, and V. E. Velculescu**. 2002. Using the transcriptome to annotate the genome. Nat. Biotechnol. **20**:508-512.

85. **Savli, H., A. Karadenizli, F. Kolayli, S. Gundes, U. Ozbek, and H. Vahaboglu**. 2003. Expression stability of six housekeeping genes: A proposal for resistance gene quantification studies of *Pseudomonas aeruginosa* by real-time quantitative RT-PCR. J. Med. Microbiol. **52**:403-408.

86. **Schena, M., D. Shalon, R. W. Davis, and P. O. Brown**. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science **270**:467-70.

87. **Sethi, S. and T. F. Murphy**. 2001. Bacterial infection in chronic obstructive pulmonary disease in 2000: a state-of-the-art review. Clin. Microbiol. Rev. **14**:336-363.

88. **Sharma, R. C. and R. T. Schimke**. 1996. Preparation of electrocompetent *E. coli* using salt-free growth medium. Biotechniques **20**:42-44.

89. **Shea, J. E. and D. W. Holden**. 2000. Signature-Tagged Mutagenesis Helps Identify Virulence Genes. ASM News **66**:15-20.

90. **Shevchenko, Y., G. G. Bouffard, Y. S. Butterfield, R. W. Blakesley, J. L. Hartley, A. C. Young, M. A. Marra, S. J. Jones, J. W. Touchman, and E. D. Green**. 2002. Systematic sequencing of cDNA clones using the transposon Tn5. Nucleic Acids Res. **30**:2469-2477.

91. **Shoemaker, D. D., D. A. Lashkari, D. Morris, M. Mittmann, and R. W. Davis**. 1996. Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy [see comments]. Nat. Genet. **14**:450-6.

92. **Smith, V., D. Botstein, and P. O. Brown**. 1995. Genetic footprinting: a genomic strategy for determining a gene's function given its sequence. Proc. Natl. Acad. Sci. U. S. A. **92**:6479-83.

93. **Smith, V., K. N. Chou, D. Lashkari, D. Botstein, and P. O. Brown**. 1996. Functional analysis of the genes of yeast chromosome V by genetic footprinting. Science **274**:2069-74.

94. **Sonnhammer, E. L., S. R. Eddy, and R. Durbin**. 1997. Pfam: a comprehensive database of protein domain families based on seed alignments. Proteins **28**:405-420.

95. **Stellwagen, A. E. and N. L. Craig**. 1997. Gain-of-function mutations in TnsC, an ATP-dependent transposition protein that activates the bacterial transposon Tn7. Genetics **145**:573-585.

96. **Swords, W. E., D. L. Chance, L. A. Cohn, J. Shao, M. A. Apicella, and A. L. Smith**. 2002. Acylation of the lipooligosaccharide of *Haemophilus influenzae* and colonization: an *htrB* mutation diminishes the colonization of human airway epithelial cells. Infect. Immun. **70**:4661-4668.

97. **Tao, H., C. Bausch, C. Richmond, F. R. Blattner, and T. Conway**. 1999. Functional genomics: expression analysis of *Escherichia coli* growing on minimal and rich media. J. Bacteriol. **181**:6425-6440.

98. **Tartof, K. D. and C. A. Hobbs**. 1987. Improved media for growing plasmid and cosmid clones, p. 12-14.

99. **Tatusov, R. L., M. Y. Galperin, D. A. Natale, and E. V. Koonin**. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. **28**:33-36.

100. **Tessier, D. C. and D. Y. Thomas**. 1996. PCR-assisted mutagenesis for site-directed insertion/deletion of large DNA segments. Methods Mol Biol **57**:229-237.

101. **Thoren, K., E. Gustafsson, A. Clevnert, T. Larsson, J. Bergstrom, and C. L. Nilsson**. 2002. Proteomic study of non-typable *Haemophilus influenzae*. J. Chromatogr. B Analyt. Technol. Biomed. Life Sci. **782**:219-226.

102. **Turk, D. C.** 1984. The pathogenicity of *Haemophilus influenzae*. J. Med. Microbiol. **18**:1-16.

103. **Velculescu, V. E., L. Zhang, B. Vogelstein, and K. W. Kinzler**. 1995. Serial analysis of gene expression. Science **270**:484-7.

104. **Vigdal, T. J., C. D. Kaufman, Z. Izsvak, D. F. Voytas, and Z. Ivics**. 2002. Common physical properties of DNA affecting target site selection of sleeping beauty and other Tc1/mariner transposable elements. J. Mol. Biol. **323**:441-452.

105. **Vinogradov, A. E.** 2003. DNA helix: the importance of being GC-rich. Nucleic Acids Res. **31**:1838-1844.

106. **Walhout, A. J., G. F. Temple, M. A. Brasch, J. L. Hartley, M. A. Lorson, H. S. van den, and M. Vidal**. 2000. GATEWAY recombinational cloning:

application to the cloning of large numbers of open reading frames or ORFeomes. Methods Enzymol. **328**:575-592.

107. **Welsh, J., K. Chada, S. S. Dalal, R. Cheng, D. Ralph, and M. McClelland**. 1992. Arbitrarily primed PCR fingerprinting of RNA. Nucleic Acids Res. **20**:4965-4970.

108. **Wilson, M., J. DeRisi, H. H. Kristensen, P. Imboden, S. Rane, P. O. Brown, and G. K. Schoolnik**. 1999. Exploring drug-induced alterations in gene expression in *Mycobacterium tuberculosis* by microarray hybridization. Proc. Natl. Acad. Sci. U. S. A. **96**:12833-12838.

109. **Winzeler, E. A. and R. W. Davis**. 1997. Functional analysis of the yeast genome. Curr. Opin. Genet. Dev. **7**:771-6.

110. **Winzeler, E. A., D. D. Shoemaker, A. Astromoff, H. Liang, K. Anderson, B. Andre, R. Bangham, R. Benito, J. D. Boeke, H. Bussey, A. M. Chu, C. Connelly, K. Davis, F. Dietrich, S. W. Dow, M. El Bakkoury, F. Foury, S. H. Friend, E. Gentalen, G. Giaever, J. H. Hegemann, T. Jones, M. Laub, H. Liao, R. W. Davis, and et al.** 1999. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. Science **285**:901-6.

111. **Zayed, H., Z. Izsvak, D. Khare, U. Heinemann, and Z. Ivics**. 2003. The DNA-bending protein HMGB1 is a cellular cofactor of Sleeping Beauty transposition. Nucleic Acids Res. **31**:2313-2322.

112. **Zhu, H., M. Bilgin, and M. Snyder**. 2003. Proteomics. Annu. Rev. Biochem. **72**:783-812.

PART II –COMPETENCE AND TRANSFORMATION

CHAPTER THREE

General Introduction: Competence and Transformation

Haemophilus influenzae is among a large number of bacterial species able to develop competence, a specific state of enhanced binding and uptake of exogenous DNA, for incorporation into the bacterial chromosome (termed transformation) (43, 51, 107, 137). Transformation has been observed in many strains of *H. influenzae* (51, 58, 107, 137) and in other members of the family *Pasteurellaceae* (59, 154). In the well-characterized *H. influenzae* strain Rd, an acapsular derivative of a type d strain (4), a state of competence develops spontaneously with the transformation frequency rising from a background level of 10^{-8} during early log-phase growth to 10^{-4} in late log-phase cells (121). Upon transfer into a starvation medium (68) or by transient anaerobiosis (52, 57), virtually 100% of cells become competent and transformation rates as high as 10^{-2} can be obtained. Transformation of *H. influenzae* has been observed *in vivo* in diffusion chambers implanted in the peritoneal cavity in rats (36) and a recent study reported horizontal transfer of *ompP2* gene sequences between *H. influenzae* strains colonizing the respiratory tract of a patient with chronic obstructive pulmonary disease (COPD) (70). The ability to exchange genetic material between colonizing organisms could result in acquisition of new virulence determinants and altered alleles contributing to avoidance of the host defenses. Alternatively, transformation ability could enhance mechanisms for adaptation to the host during colonization and may provide nucleotides as an alternate

nutrient source. Thus, competence and transformation may be an important contributor to the evolution and maintenance of virulence.

Considerable effort has been dedicated to characterize competence and transformation in *H. influenzae.* Genes previously implicated in these processes are shown in Table 3.1 together with putative or known properties of their products and selected homologs from other naturally competent organisms.

Several changes in the physiology of the cell are associated with the ability to bind and uptake DNA. Modifications of the cellular envelope include the appearance of new polypeptides, formation of membranous vesicles (transformasomes) capable of binding and protecting transforming DNA (34, 35, 37, 76, 77, 166) and altered lipooligosaccharides that may reduce electrostatic repulsion forces (167). Single-stranded gaps and tails in the chromosome appear (86, 102) and, perhaps relatedly, phage and plasmid recombination rates increase in competent cells (100, 130). Following induction of competence, the process of transformation can be divided into four distinct steps: binding, uptake, translocation and recombination.

TABLE 3.1 – Genes implicated in competence and transformation in *H. influenzae*

Region	Gene	HI #	CRE[a]	Effect[b]	Homologs[c]	Significance	Refs.
Possible direct role in transformation							
rec2	*rec2*	0061	Y	T	ComA (Ng) ComEC (Bs)	Likely cytoplasmic channel involved in translocation of the DNA into the cytosol	(12, 28, 33)
pil region	*pilA*	0299	Y	BU^{o}	PilE (Ng) ComGC/GD (Bs)	Prepilin. A possible subunit of a transformation pseudopilus	(12, 28, 42)
	pilB	0298	Y	BU^{o}	PilF (Ng) ComGA (Bs)	Traffic NTPase; responsible for pseudopilus assembly	(12, 28, 42)
	pilC	0297	Y	BU^{o}	PilG (Ng) ComGB (Bs)	Cytoplasmic membrane protein; possible involvement in pilus formation	(12, 28, 42)
	pilD	0296	Y	BU^{o}	PilD (Ng) ComC (Bs)	Prepilin peptidase; proteolytic cleavage of leader peptides from prepilins	(12, 28, 42)
com region	*comA*	0439	Y	BU†		Related to PilM? ATPase involved in Tfp formation	(149)
	comB	0438	Y	Nd		no homology outside Pasteurellaceae	
	comC	0437	Y	BU†		no homology outside Pasteurellaceae	(149)
	comD	0436	Y	Nd		no homology outside Pasteurellaceae	
	comE	0435	Y	BU	PilQ (Ng)	Outer membrane secretin	(12, 28, 149)
	comF	0434	Y	T	ComFC (Bs)	Mutation disrupts translocation; putative amidophosphoribosyltransferase	(149)
	comG	0433	Y	Nd		highly conserved, thioredoxin-like protein	
	ponA	0440	?	BU		Mild decrease; Cell wall maintenance	(149)
	comJ	0441	?	BU†		Unknown function; deletion included downstream genes	(42)
recA	*recA* (*rec-1*)	0600	N	Rec		Critical for recombination of ssDNA	(132, 168)
tfoX	*tfoX* (*sxy*)	0601	N	R		Possible regulatory role. Mutations eliminate/ overexpression enhances transformation	(159, 169)
cya	*cya*	0604	N	R		cAMP synthase/ relieved by exogenous cAMP	(41)
por	*por*	0846	N	BU	BdbDC (Bs)	Disulfide bond formation. Homologs required for pilus assembly in other organisms	(28, 147)
crp	*crp*	0957	N	R		cAMP regulatory protein/ possible direct regulator of competence	(25, 94)

TABLE 3.1 – Genes implicated in competence and transformation in *H. influenzae* (continued)

Region	Gene	HI #	CRE[a]	Effect[b]	Homologs[c]	Significance	Refs.
dpr region	*dprA*	0985	Y	T	DprA (Hp) DprA (Sp)	Possible role in protecting transforming ssDNA	(19, 81)
	dprB	0984	Y	nd		Highly conserved; unknown function. Not linked to *dprA* in other Pasteurellaceae	
	dprC	0983	Y	nd		Poorly conserved; unknown function	
comEA	*comEA*	1008	Y	nd	ComE (Ng) ComEA(Bs)	Homologs involved in DNA binding in other competent organisms	(12, 28)
comM	*comM*	1117	Y	Rec		DNA translocated normally; highly conserved predicted ATPase/Mg2+ chelatase	(60)
Likely indirect role in transformation							
pbp cluster	several	0029-39	N	BU		Peptidoglycan synthesis, maintenance	(42)
murE	*murE*	1133	N	R		Peptidoglycan synthesis. Point mutations result in hypercompetence	(92)
atp cluster	several	0478-85	N	R		ATP synthase/ relieved by exogenous cAMP	(61)
topA	*topA*	0657	N	R‡		DNA Topoisomerase I; defects alter supercoiling (pleiotropic effects on transcription)	(26, 92)
Icc	*Icc*	0399	N	R‡		cAMP phosphodiesterase	(92, 97)
trmE	*trmE*	1002	N	R‡		Possible role in tRNA modification	(92, 148)
rpoBC	*rpoBC*	0514-15	N	R‡		RNA polymerase subunits	(92, 148)
PTS cluster	*ptsI,H,crr*	1711-13	N	R		Phosphoenolpyruvate:fructose phosphotransferase system (PTS); effects cAMP levels	(96)
drug efflux region	several	0894-97	N	BU		Deletion resulted in altered growth characteristics	(42)
pgsA	*pgsA*	0123	N	BU		Synthesis of acidic phospholipids	(42)

[a] Competence regulatory element present upstream of operon. "?" indicates transcription divergent from CRE-controlled operon

[b] Effect of mutations in gene/operon. Abbreviations: BU, DNA binding/uptake defect; T, DNA translocation defect; Rec, DNA enters the cell normally, possible recombination defect; R, Regulatory effect; nd, no mutation studied/no *in silico* prediction. º Phenotype based on both polar mutations in *Hinf*, role of homologs in other bacteria. † Phenotype possibly due to polar effects. ‡ Phenotype relieved in hypercompetent *murE* double mutants.

[c] Homologs implicated in transformation of other bacteria. Abbreviations: Ng, *N. gonorrhoeae*; Bs, *B. subtilis*; Hp, *H. pylori*; *Sp*, *S. pneumoniae*.

DNA binding. The binding of exogenous DNA appears to begin as a loose binding sensitive to NaCl concentrations (38). It is followed by a tight, irreversible interaction during which the DNA cannot be removed by washing with high salt concentrations yet is still susceptible to degradation by DNase treatment (51). This indicates that there is a temporal gap between binding and uptake in *H. influenzae*. The members of the Pasteurellaceae and Neisseriaceae differ from other naturally transformable bacteria in that only donor DNA from closely related species is capable of transforming cells under normal conditions, and this bias occurs at the binding step (29). A 9-bp core (29-bp extended) consensus uptake signal sequence (USS) is responsible for the uptake bias in *H. influenzae.* The USS is identical to that of *Actinobacillus actinomycetemcomitans* but unrelated to the sequence from Neisserial species (99, 129, 154). Analysis of the completed *H. influenzae* Rd KW20 genomic sequence indicted 1465 copies of the USS are present with a mean distance of 1248-bp between sites. A significant number of these are arranged divergently, forming stem-loop structures, in intergenic regions (138). Selective pressure to maintain the consensus sites may come from a combination of their potential role as a Rho-independent transcriptional terminators and the restoring force of transformation itself due to preferential uptake of DNA carrying these sites (83). The *Haemophilus* USS is also vastly overrepresented in the genomes of the Pasteurellaceae members *Pasteurella multocida* and *Haemophilus somnus* even though natural transformation has not been observed in these two organisms (13).

DNA uptake and translocation. Following binding, the DNA duplex is taken up into a DNase-resistant form. During competence development, specialized membranous

extensions (transformasomes) appear on the surface of *H. influenzae* and *H. parainfluenzae* cells (77). These structures are visible by electron microscopy and have been shown to be composed of a lipid bilayer, competence-specific outer membrane proteins and lipooligosaccharides. The transformasomes are spherical, measure 80-100 nm in diameter and are present at an estimated 13±5 vesicles per cell (37). A pore structure with an opening of 30Å exists at the base of the transformasome (76) and the vesicles appear to be localized at points of possible fusion between the cytoplasmic and outer membranes (78). Transformation defective mutants have been isolated that spontaneously shed the vesicles and the purified vesicles interact with duplex DNA to form stable complexes resistant to DNase digestion. The uptake specificity is also associated with the transformasomes (35, 37, 75). The transfer of competent wild-type cells to normal growth media is accompanied by the shedding of vesicles into the medium and a reduction in transformation ability. These vesicles, like those of the mutants, retain the ability to bind DNA and render it DNase-resistant (37).

There is considerable evidence that the type II secretion system and type IV pili (Tfp), together referred to as the PSTC proteins (Pilus, Secretion, Twitching motility and Competence), are integral components of the transformation systems of most naturally competent bacteria (43). The exception appears to be *Helicobacter pylori*, which uses a system related to type IV secretion systems (136). The relationship between Tfp expression and competence was first identified in *N. gonorrhoeae* (140) and has since been reported in *Legionella pneumophila, Pseudomonas stutzeri, Thermus thermophilus, Synechocystis sp.* and *Ralstonia solanacearum* (48, 54, 79, 143, 164). While *H. influenzae, A. actinomycetemcomitans, Bacillus subtilis, Streptococcus pneumoniae*, and

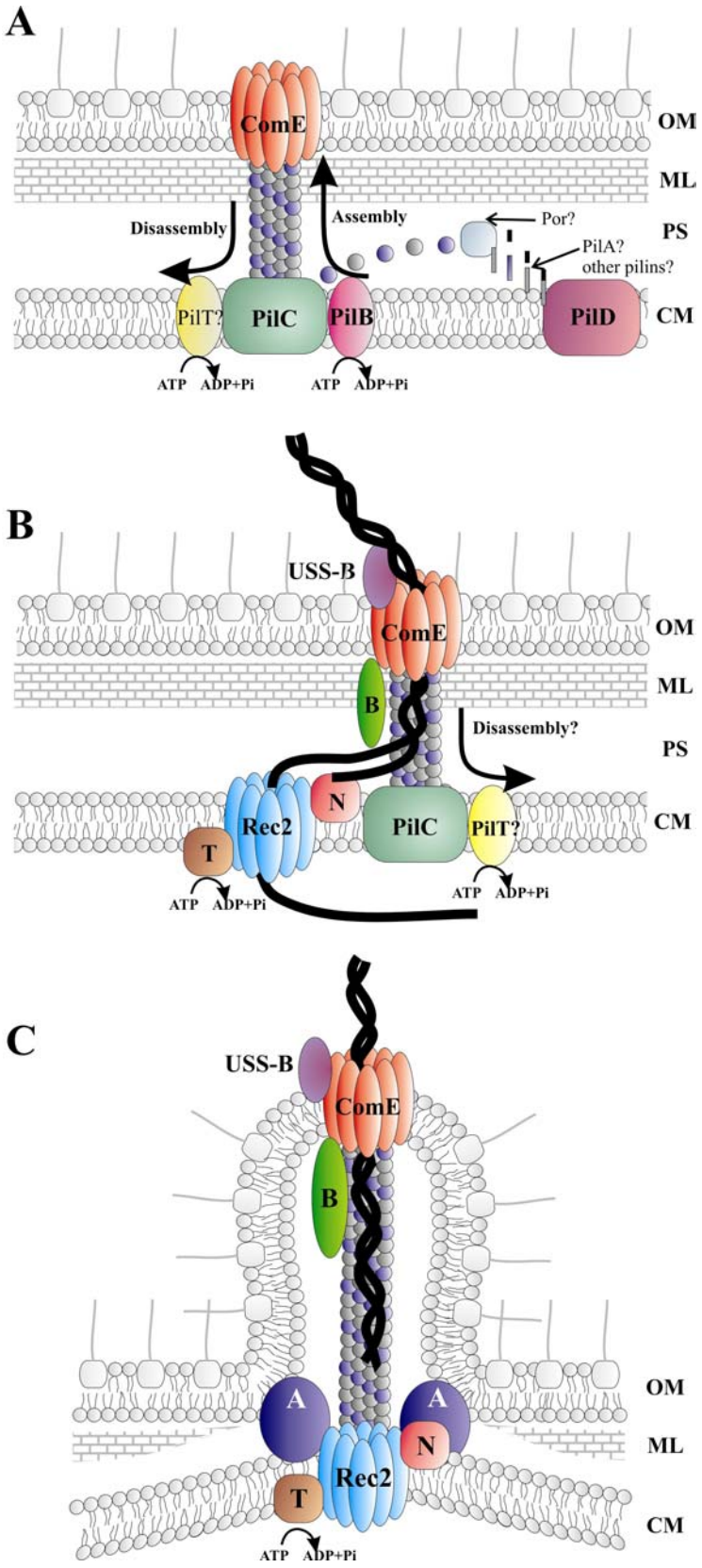
Acinetobacter sp. do not express Tfp, mutations in PSTC homologs in these organisms are associated with loss of transformability (3, 42, 111, 113, 149, 155). Recent studies in *N. gonorrhoeae* have indicated that Tfp are essential to both binding and uptake of DNA, in such manner that mutation or overexpression of various components of the system can separate the two events (2). Biosynthesis of the pilus, along with two proteins, ComP and ComE, is necessary for DNA binding. Uptake requires the action of PilT, a traffic NTPase that contributes to retraction of the pilus (161). It would be reasonable to assume that the pilus is responsible for binding DNA and retracts to bring bound DNA into the periplasm through the outer membrane pilus secretin (PilQ) or to bring the DNA into close proximity to an unidentified transport porin. In reality, while the Tfp biosynthesis machinery is required for Neisserial transformation, the pilus itself is not necessary and only small amounts of pilin are required to increase transformation levels above background (49, 91). This helps reconcile the observation that many other naturally transformable organisms, including *H. influenzae*, possess PSTC homologs but lack discernable Tfp. An alternate explanation has been proposed by Rudel *et al.* and Chen and Dubnau in which two competing structures can be formed by the PSTC machinery; intact, long Tfp and a pseudopilus-like transformation structure that incorporates additional minor pilins but does not extend through the outer membrane or a detectable distance from the cell (28, 125). In this scenario, PilT acts not as a retractor accessory but instead as an antagonist to PilF, the traffic NTPase responsible for pilus assembly, thus regulating distribution between Tfp and pseudopilus formation. Two studies of other bacteria support this hypothesis. In *N. gonorrhoeae,* overexpression of ComP, a minor pilin involved in DNA uptake, leads to increased transformation in a *pilT* mutant

background (2). In *P. stutzeri*, a modified version of the major pilin, PilAI, supports transformation but not pilus formation and suppresses the transformation deficiency in *pilT* mutants (56). This may be important as *H. influenzae* and other competent organisms do not appear to have a PilT homolog, although an analogous protein might exist (see Chapter Four discussion). Since homologs of the Neisserial PSTC proteins are necessary for transformation in *H. influenzae* and its close relative *A. actinomycetemcomitans*, it is likely that a similar structure is responsible for transformation in these organisms. The formation of a transformation pseudopilus may also account for the transformasome structures observed in competent *H. influenzae.* In *N. gonorrhoeae*, *pilQpilT* double mutants form ingrown Tfp in the periplasm that result in the formation of blebs on the outer membranes of the cells (162). A putative model of the formation of the transformation machinery in *H. influenzae*, adapted from *N. gonorrhoeae* models (12, 28), is shown in Figure 3.1A.

Once the DNA is taken into the transformasomes, it must be translocated across the murein layer and the cytoplasmic membrane. The mechanism by which this translocation occurs in transformable organisms is poorly understood. In the transformation structure proposed in Figure 3.1A, the DNA is pulled through the OM secretin and through, or alongside, the pseudopilus structure by a DNA binding protein and delivered to an inner membrane secretin (Figure 3.1B). Models of Tfp structure in *N. gonorrhoeae* and *Pseudomonas aeruginosa* propose a hollow fiber; however, the diameter (12Å) and hydrophobic nature of the cavity would likely preclude passage of DNA (47). It is possible that the additional minor pilins involved in the transformation pseudopilus may alter the structure to produce a specialized fiber with a larger diameter

designed to allow passage of transforming DNA. The observation of a 30Å pore at the base of *H. influenzae* transformasomes may represent the cavity of an altered pilus structure or the opening of a transmembrane pore. A further adaptation of the Neisserial DNA uptake model, incorporating the data on transformasomes in *H. influenzae*, is shown in Figure 3.1C.

Figure 3.1 – Model of the DNA uptake mechanism in *H. influenzae.* **A.** Following induction of competence, a dedicated transformation pseudopilus is constructed for binding and uptake of exogenous DNA. The pseudopilus is proposed to be a hollow, cylindrical fiber formed by polymerization of PilA and other pseudopilins. Prepilins are exported into the periplasm and the N-terminal signal sequences removed by the action of a prepilin peptidase (PilD). A periplasmic oxidoreductase (Por) is required for proper folding of the pilins. Assembly of the pilins into a pilus structure requires the action of a traffic NTPase (PilB). PilC homologs are required for assembly but their exact role is unknown. Tfp cross the outer membrane through a dedicated secretin (ComE homologs). While the transformation pseudopilus is not expected to extend from the cell surface, ComE is still required for DNA uptake. Disassembly of the transformation pseudopilus also requires the action of a traffic NTPase, possibly PilB or a PilT analogue. **B.** The DNA is proposed to bind to a dedicated outer membrane protein that recognizes the uptake signal sequence (USS-B). Transport into the periplasmic space likely occurs through the pore formed by ComE and may require disassembly of the pseudopilus. Other organisms require a ComEA homolog that may act to shuttle the incoming DNA to a dedicated transporter for passage across the cytoplasmic membrane (Rec2). One strand of the incoming DNA is degraded by an unknown nuclease (N) prior to, or concomitant with transport into the cytoplasm. An additional protein (T) may be required to provide energy for transportation across the membrane. **C.** Electron microscopic analysis of competent *H. influenzae* reveals the presence of membrane blebs (transformasomes) on the cell surface and apparent fusion of the outer and cytoplasmic membranes (77, 78). This may indicate a need for an additional, unrecognized protein (A) that acts to anchor the membranes together. Model of the *H. influenzae* DNA uptake mechanism adapted from an *N. gonorrhoeae* model (28). Abbrieviations: OM: Outer membrane; ML: Murein layer; PS: Periplasmic space; CM: Cytoplasmic membrane


A
ComE
OM
ML
Disassembly
Assembly
Por?
PS
PilA?
other pilins?
PilT?
PilC
PilB
PilD
CM
ATP
ADP+Pi
ATP
ADP+Pi
B
USS-B
ComE
OM
ML
B
Disassembly?
PS
N
Rec2
PilC
PilT?
CM
T
ATP
ADP+Pi
ATP
ADP+Pi
C
USS-B
ComE
B
OM
A
A
ML
N
Rec2
T
CM
ATP
ADP+Pi

The translocation of DNA across the cytoplasmic membrane is presumably through a channel formed by a conserved competence-related membrane protein. In *B. subtilis,* ComEC has 6 transmembrane regions and forms a dimer with itself. An additional membrane protein, ComFA, is required for transformation in *B. subtilis*, and resembles the DEAD family of ATP-dependent helicases (90). A third protein ComEA, contains membrane spanning and DNA binding domains and is responsible for binding DNA for uptake (72, 116). These three proteins are proposed to form a unique ABC transport system for uptake of DNA in *B. subtilis* (44). Both *H. influenzae* and *N. gonorrhoeae* have ComEA homologs (HI1008 and ComE, respectively), although they only contain the DNA-binding domain. ComE has been implicated in transformation in *N. gonorrhoeae* but is not responsible for the sequence specificity in DNA binding (30). Both organisms also have ComEC homologs (Rec2 and ComA, respectively) that are essential for transformation and lack identifiable ComFA homologs (33, 45). It is possible that an alternate ATP-dependent helicase substitutes for ComFA to facilitate the passage of DNA across the membrane.

During translocation of linear DNA, the duplex DNA is degraded by a 5'→3' nuclease, without strand bias, resulting in single-stranded DNA available for transformation (15, 53). In Gram-positive transforming organisms, there is no outer membrane to negotiate; thus, uptake and translocation are essentially accomplished in a single step. The degradation of the donor DNA is concomitant with uptake, and released nucleotides can be observed immediately upon entrance of the transforming DNA. In *H. influenzae*, degradation of the DNA occurs after uptake and the released nucleotides appear in the medium shortly thereafter (15). Thus, the degradation likely occurs in the

periplasm prior to or simultaneously with the transport across the plasma membrane and not on the cytosolic side of the membrane. The nuclease activity is blocked in *rec-2* mutants in *H. influenzae* and in *comEC* mutants in *B. subtilis* although not in *comA* mutants in *N. gonorrhoeae* (17, 27, 115). Following translocation, partial degradation of the 3'- transforming strand occurs on the cytoplasmic side of the plasma membrane regardless of homology to the chromosomal DNA. The nucleases involved in both actions have yet to be identified in *H. influenzae* or *N. gonorrhoeae*.

Several other genes have been implicated in DNA uptake or translocation in *H. influenzae.* The *com* operon includes 7 genes, *comA-G*, that have possible or known roles in transformation. ComE is a homolog of the Neisserial PilQ protein and thus is likely to be the outer membrane secretin. ComA-D show little homology to proteins outside of those from closely related organisms, but ComB and ComD have properties consistent with prepilins, *i.e.* less than 20kDa, short N-terminal leader peptide, a hydrophobic region and no sequence conservation in the C-terminal regions (28). Insertion mutations in *comA* and *comC* result in complete loss of DNA binding and uptake; however these observations may be due to polar effects since mutations in *comE* affect transformation in the same manner (149). ComF is a highly conserved protein with homology to ComFC of *B. subtilis*. Insertions in *comF* result in a phenotype of normal DNA binding and uptake but DNA remains in the periplasmic space (149). The role of the ComF homologs in translocation has yet to be determined. Dougherty and Smith isolated a deletion mutant in *comJ*, located in a putative operon divergently transcribed from the *com* operon, that resulted in the elimination of DNA binding and uptake. This deletion also included a downstream gene *ybaB* (42). Several factors have led to speculation that the *ybaB* product

may play a role in recombination; the gene is co-transcribed with *recR,* its co-arrangement with *recR* is conserved in the eubacteria, and deletion of the gene in *Streptomyces lividans* resulted in increased sensitivity to DNA-damaging agents (88, 109). In addition, *topB*, encoding Topoisomerase III, is located immediately downstream of *recR* and is likely to be cotranscribed with *ybaB* and *recR* (the predicted start codon of TopB is located 15 nucleotides from the predicted RecR stop codon). Mutations in *topA,* encoding Topoisomerase I, eliminated transformation in *H. influenzae* (26). Since the *comJ* deletion would have likely resulted in polar effects on these downstream genes, confirmation of a *comJ* role in transformation requires either complementation of the deletion mutant or a new mutation that does not affect the downstream genes.

The action of a disulfide bond oxidoreductase is required for binding and uptake of DNA. An insertion mutation in *por,* an oxidoreductase identified in the Rd KW20 genome sequence, results in the complete loss of transformation in *Haemophilus* and the elimination of competence-related changes in the outer membrane protein profile (147). These phenotypes are presumably due to the involvement of Por in stabilization and proper folding of competence proteins within the periplasm. Disulfide bond oxidoreductases have a role in stability of pilins in both Tfp and type II secretion systems (40, 118). Similarly, the oxidoreductases encoded by *bdbDC* are required for transformation in *B. subtilis* (104).

The product of the *dprA* gene, encoding a predicted Rossman-fold nucleotide binding protein, is also required for translocation of DNA in *H. influenzae* and *dprA* mutations result in a phenotype similar to that of *rec-2* mutants (81). DprA homologs are highly conserved and are required for transformation in such diverse species as *B.*

subtilis, S. pneumoniae, T. thermophilus and *H. pylori* (10, 48, 87, 108, 135). The exact role of DprA in translocation has yet to be determined.

Insertion or deletion mutagenesis has resulted in the discovery of several genes with known functions unrelated to transformation for which disruption results in defects in binding, uptake or translocation but their roles in transformation are not apparent. Insertions in genes encoding a number of penicillin binding proteins (PBPs), including PonA, RodA and Pbp2, have been shown to decrease or eliminate DNA binding and uptake, resulting in a lowered transformation frequency (42, 149). PBPs have been shown to be involved in peptidoglycan synthesis, regulation of cell division and cellular shape (152, 153). Therefore, rather than having a specific role in transformation, mutations of PBPs may disrupt cell structures and prevent the proper formation of a transformation pseudopilus across the murein layer. A deletion mutant in the multidrug-efflux region, which includes *acrA*, *acrB,* and *ftsN,* also resulted in a 20-fold decrease in DNA binding and uptake and background levels of transformation (42). Since *acrAB* mutations in *H. influenzae* result in cells hypersensitive to antibiotics, the elimination of transformation to antibiotic resistance is predicable (127). Interestingly, an insertion in *ftsN* alone resulted in background levels of DNA binding and uptake but only a 48-fold decrease in transformation frequency (42). Other mutants with similar DNA binding and uptake levels demonstrated 500-fold or greater decreases in transformation frequency. No further studies have been published to explain this phenomenon and the role that either *acrAB* or *ftsN* may play in transformation. Finally, a deletion of *pgsA* also resulted in a profound transformation defect, including a 650-fold decrease in transformation frequency and near background levels of binding and uptake (42). The product of the *pgsA* gene in *E.*

coli is involved in the synthesis of acidic phospholipids. Disruption of *pgsA* in *H. influenzae* would presumably reduce the wild-type levels (15%) of phosphatidylglycerol within the plasma membrane. The concentration of phospholipids within the membrane are not affected by the development of competence (145). Since it is unlikely to be directly involved in competence, it is possible that the alteration of phospholipid content adversely affects the efficiency of transformation in the *pgsA* mutant.

Recombination of DNA. Following uptake and translocation, the transforming DNA is incorporated into the chromosome. The recombination machinery involved is unknown but several mutants have been isolated in which translocation appears normal but the transformation frequency is reduced or eliminated. One gene that is required for transformation is *recA* (*rec-1*), and the product of this gene is the functional homolog of RecA from *E. coli* (132, 144, 168). RecA facilitates recombination by coating and protecting ssDNA from nuclease activity and by promoting strand exchange through the formation of a stable, triple-stranded intermediate between the invading strand and the duplex target DNA (82, 123). Incoming DNA in *H. influenzae rec-1* mutants is degraded in the cytoplasm and the nucleotides are recycled. Gwinn *et al.* isolated a transformation-deficient mutant with a transposon disruption that prevented transcription of *comM*. The resulting phenotypic defect was normal uptake and translocation of DNA into the cytoplasm but decreased transformation frequency and a loss of competence-related phage recombination (60). Rec2 may also play a role in transformation beyond the translocation event. Transformation mutants with a defect in *rec-2* are proficient in certain recombination events, such as between resident plasmids, but are profoundly defective in phage recombination (84, 100, 101).

The *H. influenzae* genome encodes proteins with significant homology to many recombination proteins previously identified in *E. coli*. Among these, RecO and RecR, part of the RecFOR complex, are likely candidates to be involved in transformation. RecO and RecR stabilize RecA protein filaments by preventing the end-dependent dissociation from ssDNA and also assist RecA in binding to ssDNA bound by single-strand binding protein (Ssb) (67, 133). There may also be a role for the RecBCD complex in transformation. The *in vivo* activity of RecBCD is the generation of single-stranded ends on duplex DNA through its helicase and nuclease activities. When there is an excess of Mg^{2+} over ATP, the complex behaves as a destructive nuclease, degrading both strands until it reaches a χ-site. At this point its 3'→5' nuclease activity is attenuated, its 5'→3' nuclease activity is enhanced and the complex assists in loading RecA onto the ssDNA it produces (8, 9, 39). Considering the modifications known to occur to the incoming DNA, the RecBCD complex could be involved in transformation. Evidence for the roles of these recombination complexes in transformation in other bacteria is mixed. In *N. gonorrhoeae*, mutations in the *recO* and *recR* genes have no measurable effects upon transformation while mutations in *recB, recC* and *recD* lead to a 40-fold decrease in transformation frequency (103). In *B. subtilis,* the functional homolog of the RecBCD complex is encoded by the *addA* and *addB* genes. Mutations in *addAB* reduce transformation rates in *B. subtilis* but the defect is limited to the helicase subunit of the complex and is independent of its nuclease function (6, 64). In addition, the combination of an *addAB* mutation with a mutation in the *recFOR* pathway reduces the frequency of transformation to that of a *recA* mutation, indicating that both pathways may be involved in *B. subtilis* (5). It has not been determined whether these complexes participate in

transformation in *H. influenzae* transformation. No other proteins involved in the recombination events in *H. influenzae* have been identified.

Regulation of competence development. While the basic mechanism of transformation appears to be similar between naturally-transformable organisms, the regulation of competence appears vastly different. In *S. pneumoniae* and *B. subtilis*, competence development is regulated by a quorum-sensing mechanism and the use of a competence-specific pheromone (110, 139). Upon reaching the necessary cell density, the extracellular competence peptide activates a transmembrane histidine kinase which in turn phosphorylates a regulator protein that begins a complex competence regulatory cascade (65). There is considerable polymorphism and specificity in the sensing systems between strains within both of these species. *B. subtilis* and *S. pneumoniae* do not demonstrate the uptake sequence specificity that is seen in the Pasteurellaceae or Neisseriaceae (28). Since uptake of non-species DNA carries the risk of gene disruption or production of toxic products, the polymorphism and specificity of the competence pheromone may provide a function similar to sequence specificity by limiting competence development to periods when the cell would be predominately exposed to DNA of related strains (150). Neither *H. influenzae* nor *N. gonorrhoeae* appear to be regulated in such a manner. In *N. gonorrhoeae,* Tfp production is regulated but essential to virulence (66, 85). Therefore, if competence development is to be regulated independently of Tfp production, it must occur at either the expression of competence specific components of the PSTC system or at a checkpoint governing distribution between Tfp and competence pseudopilus production. While overexpression of ComP increases transformation frequency in a dose-dependent manner, there is currently no

evidence supporting independent regulation of *comP* (the competence-specific prepilin) or any other competence-specific gene in *N. gonorrhoeae* (163). An alternate pilin subunit has been shown to have an antagonistic action towards transformability in *P. stutzeri* (55). The same effect appears to occur in *N. gonorrhoeae* with the effect of PilV on ComP (1). This action occurs in a post-transcriptional manner to affect ComP accumulation, possibly by competing for a common translocation site or by forming mixed, non-functional multimers that titrate ComP from the system.

In contrast to *N. gonorrhoeae,* the regulation of competence is well characterized in *H. influenzae.* For the last decade, much of the focus on transformation in *H. influenzae* has been on the regulation of competence development. Since Tfp are not produced in *H. influenzae*, expression of the PSTC biosynthesis machinery necessary for transformation can be tightly regulated and tied directly to the development of competence. It was noted that competence development in *H. influenzae* requires conditions in which growth of the cells is slowed but protein synthesis is allowed to continue (120, 141). Maximal levels of competence (10^{-3} to 10^{-2} transformation frequency) occur *in vitro* with a transition from rich media to a non-growth starvation media (MIV), indicating that competence development is possibly tied to the nutritional state of the cells (68). Further evidence of a nutritional role included the observation that a moderate level of competence develops spontaneously in late-log phase cultures (10^{-4} transformation frequency) but that it could be induced in early-log phase cultures by the addition of cyclic AMP (cAMP) (160). Insertions in *cya*, encoding the adenylate cyclase enzyme that produces cAMP, had no effect on growth of cells but eliminated transformation, even after transfer into MIV media (41). Addition of exogenous cAMP to

cya mutants restored transformation frequency to that of wild-type cells. Relatedly, the cAMP receptor protein (CRP), encoded by *crp*, is required for competence development and this requirement cannot be alleviated by addition of exogenous cAMP (25). In other bacteria, the cAMP-CRP complex is involved in transcriptional activation or repression of a number of genes involved in uptake and utilization of carbohydrates (23).

Other genes known to play a role in the regulation of cAMP levels in other bacteria also mediate competence development in *H. influenzae.* The phosphoenolpyruvate:carbohydrate phosphotransferase system (PTS) is involved in uptake of carbohydrates. In the presence of PTS sugars, the system represses adenylate cyclase while in the absence of the sugars it activates it. Insertion into the PTS component genes *crr* and *ptsI* led to a dramatic depression in transformation rates that was relieved by addition of exogenous cAMP (62, 96). Finally, a 3'-5' cAMP phosphodiesterase, encoded by *icc*, modulates cAMP levels in *H. influenzae*. Interestingly, while insertions in *icc* induce spontaneous competence in rich media and cAMP levels are elevated above normal, the transformation frequency in both rich and starvation media are lower than in wild-type cells, indicating a role for the phosphodiesterase in competence optimization (97).

It appears that the relief of catabolite repression has a critical role in competence development in *H. influenzae*. Most of the operons containing genes known to be directly involved in DNA uptake and transformation have a conserved 26-bp palindromic element called the competence regulatory element (CRE) directly upstream of the transcriptional start sites (46, 80) (see Figure 3.1). It has been demonstrated that *tfoX* (*sxy*) is required for transcription of transformation operons containing the CRE, and insertions into the gene

completely abolish transformation development (80, 159, 169). Additionally, the presence of *tfoX* on multicopy plasmids and a hypercompetent mutant (*sxy-1*) induces moderate transformation frequencies in noninducing conditions and fully-induced frequencies in partially-inducing conditions (121, 159, 169).

Two alternate theories exist to explain the activation of the competence regulon. In the first, the cAMP-CRP complex activates transcription of *tfoX* which in turn activates the CRE containing genes. Analysis of the protein sequence of TfoX shows no similarity to known transcriptional activators to suggest a direct role in DNA binding or transcriptional activation. Additionally, *tfoX* transcripts are prominent in exponential phase cells, and analysis of *tfoX::lacZ* fusions indicated that transcription from the *tfoX* promoter is only moderately affected by cAMP levels (14, 169). The second theory proposes that the cAMP-CRP complex is a direct activator of the competence regulon since the consensus CRE site shares some similarity to the consensus CRP recognition site (94). The cAMP-CRP complex might activate transcription of transformation genes by binding directly to the CRE, perhaps in association with a competence-specific regulator. A candidate for this second regulator is TfoX. This theory is difficult to reconcile with the observation of spontaneous competence in strains carrying multiple copies of *tfoX* or the *sxy-1* allele. In these strains, high levels of transformation occur in conditions that would presume low cAMP levels, and the addition of exogenous cAMP does not increase competence levels. A double mutant carrying the *sxy-1* mutation and the *crp* or *cya* mutations could help answer this question but this logical next step has not been reported. Previous analysis of *tfoX* mRNA indicated the presence of a stem loop structure that could play a role in either stability of the transcript or translational

regulation (14). Hypercompetent *tfoX* mutations mapped to point mutations that would theoretically destabilize the secondary structure of the *tfoX* transcript. Fusions of *β*-galactosidase with *sxy-1* (weak stem loop) and *sxy-11* (no stem loop) resulted in increased transcription and translation over fusions containing the wild-type stem loop structure. Rather than assisting in determining the roles of TfoX and the cAMP-CRP complex in competence development, this information simply indicates that regulation of the competence regulon appears more complex than anticipated.

Macfadyen and coworkers demonstrated that regulation of competence development is also dependent on the availability of certain nucleotides (95). The addition of adenosine monophosphate (AMP) or guanosine monophosphate (GMP), or their ribonucleoside equivalents, lowered the transformation frequency by several hundred-fold when induced by transfer to MIV media. Addition of AMP or GMP also caused sharp reductions of β-galactosidase activity from *comA::lacZ* and *rec-2::lacZ* fusions. No transformation effects were observed with the addition of NTPs, dNTPs, dNMPs, adenine, guanine or pyrimidine-based NMPs. The effects of AMP and GMP did not appear to be mediated by the PTS system or the stringent response. The addition of cAMP partially overcame AMP-inhibited transformation and fully restored GMP-inhibited transformation, but required cAMP levels in excess of that necessary for restoration of transformation in *cya* mutants. While this may indicate that the transformation defect observed is not due to lowered cAMP levels, the intracellular levels of cAMP was not measured and therefore it cannot be ruled out. An alternative explanation is that the competence regulon is also regulated by the PurR repressor, which allows transcription only when the purine pools are depleted (124, 165). The presence of

putative PurR binding sites upstream of *rec-2* and *dprA* would appear to support this theory. Joint regulation of the competence regulon is consistent with regulation of other operons controlled by catabolite repression (23, 95). In those situations, the increase in cAMP indicates a lack of preferred energy sources and the cAMP-CRP complex acts as a global regulator. Activation of individual operons is through a specific signal that indicates the utility of those genes to supply the cell's needs. If this is the case, this would indicate a role for transformation in *H. influenzae*, *i.e.* nutrient acquisition, which appears to be distinct from that observed in other competent bacteria. However, unpublished communication from Redfield and colleagues indicates that disruption of *purR*, while relieving nucleotide repression of transcription from *rec-2*, has no overall effect on repression of competence development caused by the addition of nucleotides (122).

Several transformation defective mutants have been identified that carry mutations in genes that, based upon the known functions of their products, are more likely to play a coincidental role in competence development. Most prominent among these is *topA,* encoding DNA topoisomerase I (TopI). It would be expected that DNA topoisomerases would be involved in recombination of transforming DNA by relieving any supercoiling that would occur from the process. This would presume that any transformation defect in a *topI* mutant would be confined to decreasing transformation efficiency and not affecting the transport of DNA into the cell. However, insertions into *topA* lower DNA binding, uptake and transformation frequency to background levels (26, 148). An increase in test plasmid supercoiling has been observed *in vivo* in the *topI* mutants, along with increased ultraviolet sensitivity and decreased growth rate. The probable result of *topA* mutations is to lower transcription levels of competence genes

due to topology constraints of the chromosome, as has been observed for many genes in other bacteria (69, 142). Further evidence to support this conclusion is the finding that other promoters regulated by the cAMP-CRP complex are sensitive to DNA superhelicity (21). Examination of the transcriptional profiles of competence genes in the *topA* mutant would be enlightening.

Insertions into *atpA*, encoding the α subunit of the F_1 domain of the ATP synthase, and *atpB,* encoding the *a* subunit of the F_0 domain, resulted in moderate decreases in transformability. Beta-galactosidase production from *comA::lacZ* and *rec-2::lacZ* fusions in the *atpA* mutant was decreased significantly compared to cells containing an intact *atpA* (61). The loss of a functional ATP synthase leads to decreased ATP levels in *E. coli* (73), which may cause increased AMP levels in the cell, a condition shown to negatively impact competence development (95). Transposon insertions into *trmE* and in the intergenic region between *rpoB* and *rpoC* also reduce DNA binding and uptake to near background levels and result in a 1000-fold decrease in transformation frequency (92, 148). Both insertions appear to affect competence development with decreased β-galactosidase activity from *comA::lacZ* and *rec-2::lacZ* (122). The genes *rpoB* and *rpoC* encode the β and β' subunits of the RNA polymerase (11). The insertion in the *rpoBC* mutant would presumably limit *rpoC* expression and would lead to decreased global transcription. The *trmE* gene encodes a GTPase that has been shown to be involved in the biosynthesis of the nucleoside 5-methylaminomethyl-2-thiouridine located in the wobble position of some tRNAs (24, 63). The frequency of translational frameshifts at codons AAA and AAG (encoding lysine) increases in *trmE* mutants (151).

It is unknown how the frameshift phenotype in *trmE* mutants specifically causes the observed transformation defects.

Finally, several point mutations in the essential peptidoglycan synthesis gene *murE* result in dramatically increased levels of competence in rich media (92). MurE is responsible for contributing *meso*-2,6-diaminopimelate to the peptide side chain of UDP-*N*-acetylmuramic acid (98). These point mutations appear to have no effect on peptidoglycan synthesis and the cells appear normal. β-galactosidase activity in strains containing *comA::lacZ* and *rec-2::lacZ* fusions is elevated in the *murE* hypercompetent mutants, and the effects of the *murE* mutations are not mediated by increased expression of *tfoX* (92). Perhaps most interestingly, the *murE* point mutations are able to elevate competence levels to near wild-type when combined with *thdF*, *topA*, *icc*, and *rpoBC* mutations that reduce but do not eliminate transformation. *murE* point mutations are unable to complement *cya*, *crp*, *tfoX*, *rec-2*, *dprA*, or *comE* mutations, each of which completely abolishes transformation. The inability to complement mutations in genes believed to be directly involved in competence induction and DNA uptake indicates that both regulation and uptake are mediated by the normal pathways in the *murE* mutants. The means by which the *murE* mutants affect competence development is not apparent, and speculation that it was due to changes in peptidoglycan recycling has been proven incorrect (92).

Our understanding of the processes involved in the transformation machinery and the development of competence in *H. influenzae* remains incomplete. Similarities between transformation systems in other bacterial species can provide additional information to understand transformation in *H. influenzae.* Even so, it is likely that

previously undiscovered factors specific to *H. influenzae* transformation exist. Furthermore, the regulation of competence in the Pasteurellaceae appears to be unique and warrants further investigation. To these ends, the transposon mutagenesis libraries created in Chapter Two were used to screen for novel transformation-related genes in Rd KW20. In addition, *in silico* predictions and comparative genomics were combined with transcriptional and mutation analyses to identify additional factors that may effect transformation in this organism. The results of these studies are presented in Chapter Four.

CHAPTER FOUR

Competence and Transformation Studies in *Haemophilus influenzae*

ABSTRACT

Haemophilus influenzae is one of a growing number of bacteria identified that possess the natural ability to uptake exogenous DNA for potential genomic transformation. Several operons involved in transformation in this organism have been described. These operons are characterized by a conserved 22-bp regulatory element upstream of the first gene and are induced coincident with transfer from rich to nutrient-depleted media. The previously identified operons are comprised of genes encoding proteins that include components of the type II secretion system and type IV pili, shown to be essential for transformation in other bacteria, and encoding other proteins previously identified as required for transformation in *H. influenzae*.

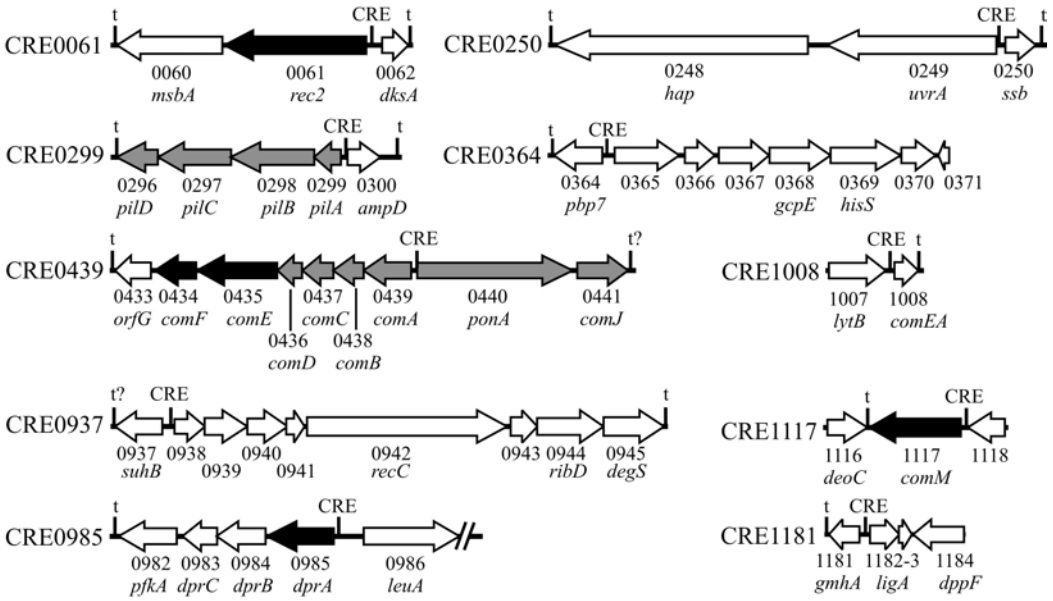
In this study, three novel competence operons were identified by comparative genomics and transcriptional analysis. These operons have been further characterized by construction of null mutants and examination of the resulting transformation phenotypes. The putative protein encoded by HI0366 was shown to be essential for DNA uptake, but not binding, and demonstrates homology to a protein shown to be required for pilus biogenesis and twitching motility in *Pseudomonas aeruginosa.* An insertion in HI0939 abolished both DNA binding and uptake. The predicted HI0939 product shares characteristics with PulJ, a pseudopilin involved in pullulanase export in *Klebsiella oxytoca.* In addition to the three competence-regulated operons, a Tn5 insertion in the conserved gene HI1161 was shown to negatively impact transformation by affecting transcription of competence-specific genes.

INTRODUCTION

Most of the operons containing genes known to be directly involved in DNA uptake and transformation in *H. influenzae* have a conserved 22-bp palindromic element, the competence regulatory element (CRE), directly upstream of the transcriptional start sites (80, 94). Nine putative CRE sites have been previously identified in the *H. influenzae* Rd KW20 genome (Figure 4.1). In addition, a site identified as a putative catabolite repressor binding (CRP) site (94), located upstream of HI0937, is nearly identical to the consensus CRE site. Mutational analyses of genes in five of these operons have confirmed their participation in transformation in *H. influenzae* (33, 42, 60, 81, 149). Competence-induced transcription of several of these operons has been demonstrated using *lacZ* fusions or northern blot analyses (60, 61, 80). While eight of the CRE sites are located between divergently transcribed operons, bidirectional control of transcription from a CRE site has not been reported.

In silico, transcriptional and mutational analyses were applied to characterize three of the previously unexamined CRE regions (CRE0364, CRE0937 and CRE1181) to determine their potential role in transformation in *H. influenzae.* Each of these CRE sites is located between divergently oriented operons. In addition, directed mutagenesis of genes controlled by these CRE sites was performed to assay involvement in competence and transformation in *H. influenzae*. Finally, the Tn5 insertions created and mapped during a global transposon mutagenesis study (Chapter Two) were transformed into *H. influenzae* Rd KW20 and screened for defects in transformation ability.

A



B

CRE	Predicted Target(s)	Putative CRE Sequence	Transformation Association
Consensus	-	**TTTTGCGATC**NN**GATCGCAAAA**	-
CRE0061	*rec-2*	**TTTTGCGATC**CAT**ATCG**T**AAAA**	YES (33)
CRE0250	*ssb*	**TTTTGCGATC**ATT**ATCGCA**T**A**T	ND
CRE0299	*pilA-D*	**TTTTGCGATC**AG**GATCGCA**G**AA**	YES (42)
CRE0364	*pbp7*	**TTTTGCGATC**TA**GATCGCAAA**T	ND
CRE0439	*comA-G*	**TTTTGCGATC**CG**CATCG**T**AAAA**	YES (149)
CRE0937	*suhB*	**TTT**M**GCGATC**TGT**ATCGCAAA**G	ND
CRE0985	*dprA-C*	**TTTTGCGATC**TGC**ATCGCAAAA**	YES (81)
CRE1008	*comE1*	**TTTTGCGATC**GA**GATCGCAAAA**	ND
CRE1117	*comM*	**TTTTGCGATC**TA**GATCGCAAAA**	YES (60)
CRE1181	*gmhA*	**TTTTGCGAT**TTA**GATCG**A**AAAA**	ND

Figure 4.1 - Putative CRE regions in *H. influenzae* Rd KW20. A. Organization of CDSs contiguous to the putative CRE elements in the Rd KW20 genome. CRE numbers assigned are based upon the original designation of the predicted targeted gene (94). Genes shaded black indicate biological data confirming their role in transformation in *H. influenzae*. Genes shaded grey indicate mutations affecting transformation exist in the operon but polar effects prevent individual characterization. CRE0937 was originally designated as a CRP binding site (94). Putative terminators are designated "t". **B.** Comparison of the putative CRE sequences in the Rd KW20 genome and association with transformation in *H. influenzae*. "ND" indicates role undetermined.

MATERIALS AND METHODS

Materials and supplies Deoxynucleotides, T4 DNA ligase and all restriction enzymes were acquired from New England Biolabs (NEB), Beverly, Mass, unless otherwise noted. DNeasy Tissue kit, RNeasy-mini kit, RNA Protect reagent, and oligonucleotide primers (Table 4.2) were obtained from Qiagen, Valencia, Calif. Wizard Plus miniprep DNA Purification System was purchased from Promega, Madison, Wisc. SYBR Green PCR Master Mix, Taqman Reverse Transcription reagents and quantitative PCR supplies were purchased from AppliedBiosystems, Foster City, Calif. Bacterial growth media was obtained from BBL/Difco (Becton Dickenson, Sparks, Md.). T4 DNA polymerase and RQ1 RNase-free DNase were acquired from Fisher Bioreagents, Pittsburgh, Pa. Taq polymerase and Quick-spin Sephadex G-50 columns were purchased from Roche Applied Science, Indianapolis, Ind. Radiochemicals were obtained from Amersham Biosciences, Piscataway, N.J. All amino acids were purchased from Calbiochem, San Diego, Calif. All other chemicals, unless noted, were obtained from Sigma-Aldrich, Saint Louis, Mo.

Bacterial growth conditions. *H. influenzae* strains were cultured at 37°C on chocolate II agar or brain heart infusion (BHI) agar supplemented with 10 μg/ml of hemin and 10 μg/ml of β-nicotinamide adenine dinucleotide (sBHI). Liquid cultures were grown in sBHI broth. *Escherichia coli* strains were cultured at 37°C in Luria-Burtani (LB) broth (1% Tryptone, 0.5% Yeast Extract, and 170 mM NaCl) or LB agar (LB broth containing 1.5% Bacto-agar). See Table 4.1 for antibiotic concentrations used. LB agar was supplemented with 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) at 40 μg/ml when appropriate.

TABLE 4.1– Strains and plasmids used in this work

Strain or plasmid	Relevant characteristics	Source or ref.
Strains		
E. coli [a]		
TOP10	F⁻ *mcrA* Δ(*mrr-hsdRMS-mcrBC*) Φ80*lacZ*Δ*M15*Δ*lacX74 recA1 deoR araD139* Δ(*ara-leu*)*7397 galU galK rpsL* (Str^r) *endA1 nupG*	Invitrogen
DH5α	F⁻ Φ80*lacZ*Δ*M15* Δ(*lacIZYA-argF*)*U169 recA1 deoR hsdR17*(r_k^-, m_k^+) *phoA supE44 thi-1 gyrA96 relA1 endA1*	Invitrogen
H. influenzae [b]		
Rd KW20	Capsule deficient type d derivative, sequenced strain (46)	ATCC (4)
MAP9	Rd strain carrying multiple antibiotic resistance markers (Nov^r, Str^r)	D McCarthy (101)
Rec-1	Rd *rec1*	J Setlow (16, 131)
TMV15	Rd KW20, HI0365::SPEC Sp^r	This study
TMV19	Rd KW20, ΔHI0366 Sp^r	This study
TMV23	Rd KW20, Δ*ligA* Sp^r	This study
TMV24	Rd KW20, HI0939::SPEC Sp^r	This study
TMV30	TMV19, ΔHI0366 Sp^r, pTV30 Cm^r	This study
TMV1767	Rd KW20, HI1159m::Tn5 Km^r	This study
TMV1778	Rd KW20, HI1160(*hemH*)::Tn5 Km^r	This study
TMV1874-1	Rd KW20, HI1161::Tn5 Km^r	This study
TMV1874-2	Rd KW20 transformed with PCR of HI1161 locus from TMV1874-1	This study
Plasmids		
pCR2.1-TOPO	3.9-kbp TA cloning vector, P_{lac} *lacZα* Ap^r Km^r, f1 ori, ColE1 ori	Invitrogen
pSPECR	pCR-Blunt containing a 1.2-kbp Sp^r cassette (SPEC)	(158)
pUC18N	pUC18 with a *Not* I linker added at the *Hin*d III site	(146)
pSU2718	2.3-kbp *E. coli-H. influenzae* shuttle vector, P_{lac} *lacZα* Cm^r, p15a ori	
pASC1767	pASC18 carrying a 3.8-kbp *Pvu*II fragment from Rd KW20, HI1159m::Tn5	Chapter 2
pASC1778	pASC18 carrying a 3.8-kbp *Pvu*II fragment from Rd KW20, HI1160(*hemH*)::Tn5	Chapter 2
pASC1874	pASC18 carrying a 3.8-kbp *Pvu*II fragment from Rd Kw20, HI1161::Tn5	Chapter 2
pTV05	pCR2.1-TOPO carrying a 2.6-kbp PCR product from Rd KW20 (CDS HI0937-HI0940)	This study
pTV10	pCR2.1-TOPO carrying a 4.0-kbp PCR product from Rd KW20 (CDS HI0364-HI0367)	This study
pTV15	pTV10 with SPEC cassette cloned into *Swa*I site in HI0365	This study
pTV16	pCR2.1-TOPO carrying a 1.4-kbp PCR product from Rd KW20 as upstream flanking region (HI0366)	This study
pTV17	pCR2.1-TOPO carrying a 1.0-kbp PCR product from Rd KW20 as downstream flanking region (HI0366)	This study
pTV18	pUC18N containing the *Eco*RI-*Bam*HI fragment from pTV16 and *Hin*dIII-*Bam*HI fragment from pTV17	This study
pTV19	pTV18 with SPEC cassette cloned into the *Bam*HI site to create a deletion of HI0366	This study
pTV20	pCR2.1-TOPO carrying a 0.6-kbp PCR product from Rd KW20 as upstream flanking region (HI1182)	This study
pTV21	pCR2.1-TOPO carrying a 0.5-kbp PCR product from Rd KW20 as downstream flanking region (HI1183)	This study
pTV22	pUC18N containing the *Eco*RI-*Bam*HI fragment from pTV20 and *Hin*dIII-*Bam*HI fragment from pTV21	This study

TABLE 4.1 (continued) – Strains and plasmids used in this work

Strain or plasmid	Relevant characteristics	Source or ref.
pTV23	pTV22 with SPEC cassette cloned into Bam HI site to create a deletion of HI1182 (*ligA*)	This study
pTV24	pTV05 with SPEC cassette cloned into *Swa*I site in HI0939	This study
pTV25	pCR2.1-TOPO carrying a 1.5-kbp PCR product from Rd KW20 (CDSs HI0938 and HI0939 and associated CRE site)	This study
pTV26	pCR2.1-TOPO carrying a 2.4-kbp PCR product from Rd KW20 (CDSs HI0938-0941 and associated CRE site)	This study
pTV27	Shuttle plasmid containing the *Hin*dIII-*Bam*HI fragment from pTV25 cloned into the *Hin*dIII-*Bam*HI site of pSU2718	This study
pTV28	Shuttle plasmid containing the *Hin*dIII-*Bam*HI fragment from pTV26 cloned into the *Hin*dIII-*Bam*HI site of pSU2718	This study
pTV29	pCR2.1-TOPO carrying a 2.1-kbp fragment from Rd KW20 (CDS HI0365 and HI0366 and associated CRE site)	This study
pTV30	Shuttle plasmid containing the *Hin*dIII-*Bam*HI fragment from pTV29 cloned into the *Hin*dIII-*Bam*HI site of pSU2718	This study
pTV31	pCR2.1-TOPO carrying an 867-bp PCR product from Rd KW20 (CDSs HI1182 and HI1183)	This study

[a] Antibiotic concentrations used for selection of *E. coli*: spectinomycin (Sp) 150 μg/ml, ampicillin (Ap) 100 μg/ml, kanamycin (Km) 50 μg/ml, chloramphenicol (Cm) 50 μg/ml

[b] Antibiotic concentrations used for selection of *H. influenzae*: spectinomycin (Sp) 200 μg/ml, novobiocin (Nov) 2.5 μg/ml, kanamycin (Km) 20 μg/ml, chloramphenicol (Cm) 2 μg/ml, streptomycin (Str) 250μg/ml.

TABLE 4.2 – Sequences of oligonucleotides used in this work

Primer name	Sequence (5' to 3')
1182-UP-A	GAATTCATCTTTGCCTGTCATC
1183-UP-B	GGATCCCTAAGAATTTACGT
1183-DN-A	GGATCCCCTATTTATAAATCTGAC
1183-DN-B	AAGCTTCAACGTATTGCCAT
1182-F	ATCTTACGTAAATTCTTAG
1183-R	GTCAGATTTATAAATAGG
0366-UP-A	GAATTCCGCTAGATTATGC
0366-UP-B	GGATCCTTTTAATTGGGCAGGAAG
0366-DN-A	GGATCCAAATCAAGTATTGGAC
0366-DN-B	AAGCTTTGTAAGTGATGCGC
0366-R	TTCGGCGAATTAAGTGCTAATTC
0364CloneF	AAAGTCACTGTAGAACAGCCTGC
0364CloneR	AATAGTTGGCTGAAAAGCTGAC
0938CloneF	GATTAAATTCAGCACCTAAACAAGG
0938CloneR	AAAGTGCGGTCGAAAATAGG
0365-6F	CGCAAAAGAAGGACACAAAG
0365-6R	CGGGATCCAATACTTGATTTGGC
0938-41F	AACATTGGATCCATATTTCCACCA
0938-39R	CCCAAGCTTAGTCAGCGTGATAATGCC
0938-41R	CCCAAGCTTACAACGCCATTTTACTGAG
1161CloneF	TTGCCGCAGTCAAGTAAAAC
1161CloneR	GGATTGTTGCAGTATTTCAGTAAGA
QPCR-*tfoX*-F	GCTTTTGGCGAGGATTGGAT
QPCR-*tfoX*-R	TCAGCTAAAGCAACCGAAACC
QPCR-rec2-F	ACGCTTATCGCCACAGCAA
QPCR-rec2-R	AGGCACCTCTTTCGCTTTCC
QPCR-comA-F	GCACTTTACAAATCGGCATTCA
QPCR-comA-R	TGTGGCTGTTCGAGATCATCA
QPCR-0364-F2	TCGAATTAAAGGCACTGGAACA
QPCR-0364-R2	GGGCGGCATAGTTATCAGAATG
QPCR-0366-F2	GCGGTTATTTTCCCTTTCATTTT
QPCR-0366-R2	GTTCCACACGCGCTTTAGC
QPCR-0368-F	CGTGCCGTTGTTGATTGTG
QPCR-0368-R	GGTTCGCCATATTTTTCTTGCA
QPCR-0937-F3	GGATTATTGATCCGCTAGATGGTACT
QPCR-0937-R3	CCCGACTTCAGTGCGATTTT
QPCR-0938-F	TTTGCAGTACCATTATGGAAAACC
QPCR-0938-R	TCTGCCCGAGCCTGAATTT
QPCR-0939-F	ACAAACGCAAAATCAACACATGT
QPCR-0939-R	GAAATCCTAATCGGCGAAGATCT
QPCR-0942-F	AAATTTCCTATGCCAGCCAGTTT
QPCR-0942-R	AAACGCCACATCATTGAATCTTT
QPCR-1181-F	GGCATTATTAATTTCGAATAGCTTCA
QPCR-1181-R	CGTTAATTCTTCTGCAAAATGCA
QPCR-1182-F	GGGTTGGGTAATGTCTGAAAAGTT
QPCR-1182-R	AATAAGCTGGCGGCGATAAA

Nucleotide sequencing. Automated nucleotide sequencing was performed by the Recombinant DNA/Protein Resource Facility at Oklahoma State University, Stillwater, Okla.

Computational analysis. Partial annotation of incomplete genomic sequences was accomplished using Artemis (version 5) (126) and the BLAST algorithms (NCBI) (7, 50). The presence of export signal sequences was detected using SignalP 2.0 (http://www.cbs.dtu.dk/services/SignalP-2.0/) (106). The likelihood of a signal peptide cleavage site in a particular protein was scored as a positive or negative on four different properties using the neural-network (NN) model and as a secretory or non-secretory protein in the Hidden-Markov model (HMM). Proteins scoring at least 2 positive results in the NN model and identified as a probable secreted protein using the HMM method were classified as putative exported proteins for this study. Other DNA and protein sequence analyses were performed using VectorNTI Advance (version 9.0) (Informax, Frederick, Md.).

Isolation of plasmid and chromosomal DNA. Plasmids were purified from *E. coli* using the Wizard Plus miniprep kit using the manufacturer's protocol. Chromosomal DNA from *E. coli* and *H. influenzae* strains was prepared with the DNeasy Tissue kit using the manufacturer's protocol for isolation of DNA from Gram-negative bacteria.

Polymerase Chain Reaction (PCR) protocol. Each 50 µl reaction mixture was composed of 1X PCR reaction buffer (10 mM Tris-HCl, 1.5 mM MgCl2, 50 mM KCl, pH 8.3), 25 pmoles of each primer, 0.5 µg template DNA, 200 µM dNTP mix (50 µM each of dATP, dCTP, dGTP, dTTP), and 2.5 units Taq DNA Polymerase. Thirty cycles of PCR were performed (each cycle consisted of denaturation at 95°C for 1 minute,

annealing at 55°C for 1 minute, and extension at 72°C for 1 minute/kbp expected product size) followed by extension at 72°C for 10 minutes.

Cloning and ligation reactions. Restriction-digested vectors, fragments and PCR products were resolved by agarose gel electrophoresis in 1 x TBE buffer (89 mM Tris base, 89 mM boric acid, 2 mM EDTA, pH 8.0), excised from the gel and purified using the Geneclean III kit (BIO101, Carlsbad, Calif.) prior to use in cloning and building of mutagenic constructs. PCR products were cloned into pCR2.1-TOPO (Invitrogen) following the manufacturer's directions. The use of this topoisomerase-mediated TA cloning system eliminated the necessity of a standard ligation step for the cloning of PCR products. Other ligation reactions were performed overnight at 16°C in a 15 µl reaction mixture containing an approximate 1:4 molar ratio of vector to insert DNA, 800 units T4 DNA Ligase and 1X T4 DNA Ligase buffer (50 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 10 mM dithiothreitol, 1mM ATP, and 25 µg/ml bovine serum albumin).

Transformation of plasmids into *E. coli*. Plasmid constructs were transferred into electrocompetent or chemically-induced competent *E. coli*. Electrocompetent DH5α was produced using the method of Sharma and Schimke (134). To introduce the plasmids, 100 µl of prepared cells were thawed on ice, mixed with 1-2 µl of the plasmid solution, and the mixture was added to an ice-cold electroporation cuvette. Electroporation was performed in an Eppendorf 2510 at 15 kV/cm. One ml of SOC media (2% Tryptone, 0.5% Yeast Extract, 10 mM NaCl, 1 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 0.2% glucose, pH 7.0) was immediately added and the mixture was transferred into a sterile 15 ml tube. After incubation with shaking for 1 hour at 37°C to allow expression of the antibiotic markers, aliquots of the mixture were plated on LB

agar plates containing the appropriate antibiotic for selection of the correct construct. Transformation using chemically competent cells was performed using commercially prepared *E. coli* TOP10 cells (Invitrogen, Carlsbad, Calif.) following the manufacturer's directions and expression and plating were performed as described above.

Transformation of contructs into *H. influenzae.* Mutagenic constructs were transformed into *H. influenzae* Rd KW20 by two methods. The first method was an adaptation of the static aerobic transformation method of Gromkova *et al*. (58). The *H. influenzae*-specific DNA was cleaved from the vector backbone using an appropriate restriction endonuclease, and 5 µl of the digestion mixture (approximately 0.2 µg total DNA) was added to a well of a microtiter plate. Exponential-phase Rd KW20 cells were diluted 100-fold in fresh sBHI and 200 µl added to the well containing the mutagenic DNA. The plate was incubated statically overnight at 30°C to allow growth and development of competence. Following incubation, 5 µl from the well were spotted on sBHI agar containing the appropriate antibiotic for selection of transformants and the plate incubated overnight at 37°C. When transformants were not recovered with the static aerobic method, the more efficient, but laborious, MIV method was used (68). A culture of Rd KW20 was grown to an OD_{600} 0.3 and pelleted by centrifugation at 4000 x g. The cells were washed twice in prewarmed MIV media and resuspended in a volume of MIV equal to the original culture volume. Competence was induced by incubation with shaking at 37°C for 100 minutes. Mutagenic DNA was prepared as described above and 5 µl of the digestion mixture were added to 200 µl of competent cells. Following 20 minutes incubation at 37°C, 2 volumes of sBHI was added and the mixture allowed to incubate for an additional 2 hours at 37°C to allow for expression of the antibiotic

resistance gene. As above, 5 µl of the mixture was spotted on sBHI agar containing the appropriate antibiotic.

Two methods were employed to establish plasmids in *H. influenzae* strains. In the first, cells were made competent using the MIV procedure. Closed circular plasmid constructs (approximately 0.5 µg) were added to 1 ml of competent cells, and the mixture was allowed to incubate for 30 minutes at 37°C. Sterile glycerol was added to the incubated cells to a final concentration of 30% and the mixture was incubated an additional 10 minutes at room temperature (112). Two milliliters of sBHI were added to the cells and the mixture was allowed to incubate for 2 hours at 37°C to allow expression of the plasmid-borne antibiotic marker. Plasmids were also introduced into *H. influenzae* strains by electroporation using a modification of the method of Mitchell *et al.*(105). In short, cultures of *H. influenzae* were grown to an OD_{600} 0.4 then chilled on ice for 30 minutes. Cells were collected by centrifugation at 4000 x g for 5 minutes and washed 5 times with ice-cold PSG buffer (15% glycerol, 272 mM sucrose, 2.43 mM K_2HPO_4, 0.57 mM KH_2PO_4, pH 7.4). Cells were pelleted after each wash by centrifugation at 4000 x g for 10 minutes. The first two washes were performed in a volume equal to the original culture volume. Each of the three successive washes was performed in an amount of PSG buffer equal to half of the previous wash. Following the final wash, the cells were resuspended in a volume of PSG buffer equal to $1/100^{th}$ the original culture volume. In order to introduce plasmids, 0.5 µg of closed circular plasmid were added to 75 µl of freshly prepared electrocompetent cells and the mixture was electroporated at 14 kV/cm. One ml of sBHI was added to the electroporated cells, and the mixture was allowed to

incubate for 2 hours at 37°C to allow expression of the plasmid-borne antibiotic marker prior to plating on sBHI containing chloramphenicol.

Screening for transformation mutants. The transposon-generated insertion constructs described in Chapter Two were digested with *Asc*I and transformed into *H. influenzae* Rd KW20 by either the static aerobic or MIV method. The kanamycin- or chloramphenicol-resistant transformants were analyzed for their ability to be transformed to a second antibiotic-resistant phenotype using the static aerobic method. Single colonies from each transposon mutant were transferred to a well on a microtiter plate containing 200 µl sBHI broth and 0.2 µg of MAP9 chromosomal DNA and the plates were allowed to incubate overnight at 30°C to allow development of competence and transformation. Five microliters from each well were spotted onto a sBHI agar plate containing 250 µg/ml streptomycin or 2.5 µg/ml novobiocin (test) or onto sBHI (control) and the plates were incubated overnight at 37°C. Transformation proficient strains appear as numerous colonies at the spotting site. Those strains in which the growth on the control plate was normal but growth on the plates containing antibiotics was diminished or absent were designated as putative transformation mutants. These were further analyzed by repeating the static aerobic transformation and verified by the MIV method to determine if the transformation defect was reproducible.

Gene expression during competence development. The kinetics of competence induction in MIV media was correlated with transcriptional analysis of potential CRE-controlled operons using quantitative real-time PCR (Q-PCR). A 100 ml culture of *H. influenzae* Rd KW20 was grown in sBHI to an OD_{600} 0.3. A control was obtained by removing a 1 ml sample and mixing with 2 ml of RNA Protect (Qiagen) to stabilize the

RNA. The remainder of the culture was collected by centrifugation at 3000 x g, washed once with prewarmed MIV media, resuspended in 100 ml MIV media, and incubated with vigorous aeration at 37°C. One ml samples for analysis of RNA profiles were taken at 0, 20, 40, 60, 80, 100, 120, 140, and 160 minutes and mixed with 2 ml of RNA protect. The samples were incubated for 10 minutes at room temperature, followed by centrifugation at 14000 x g to pellet the cells. The supernatant was aspirated and the pellets were frozen at -20°C until processing. RNA from each sample was isolated using the RNeasy mini kit (Qiagen) as directed by the manufacturer and resuspended in 30 µl of RNase-free water. Residual chromosomal DNA was removed by digestion with amplification grade DNase I (Invitrogen) as directed by the manufacturer. The RNA samples were used to prepare cDNA in a 20 µl reaction containing 7 µl template RNA, 5.5 mM $MgCl_2$, 500 µM each dNTP (dATP, dCTP, dGTP, dTTP), 1 x RT reaction buffer, 0.08 units RNase Inhibitor, 2.5 µM random hexamers, and 25 units MultiScribe Reverse Transcriptase (AppliedBiosystems, Foster City, Calif.). The synthesis reaction was incubated at 25°C for 10 minutes followed by 48°C for 30 minutes. The reaction was terminated by heating at 95°C for 5 minutes. Prior to analysis, the cDNA was diluted by addition of 180 µl RNase-free water. Q- PCR was utilized to examine transcription of 16s rRNA (normalizer), known competence-related genes (*tfoX*, *comA*, and *rec-2*) and possible competence-regulated genes identified in this study (HI0364, HI0366, HI0368, HI0937, HI0938, HI0939, HI0942, HI1181 and HI1182).

Quantitative PCR. Gene-specific oligonucleotide primers were designed using Primer Express 2.0 (Applied Biosystems). Primers were tested to determine amplification specificity, efficiency, and linearity of the amplification to RNA

concentration as described by the manufacturer. A typical 25 µl reaction contained 12.5 µl of SYBR Green Master Mix, 250 nM each primer, and 5 µl of cDNA sample. Quantification reactions for each gene at each timepoint were performed in triplicate and normalized to concurrently run 16s rRNA levels from the same sample. Relative quantification of gene expression was determined using the $2^{-\Delta\Delta Ct}$ method of Livak and Schmittgen where $\Delta\Delta C_t = (C_{t,Target} - C_{t,16s})_{Time\ x} - (C_{t,Target} - C_{t,16s})_{Control}$ (89).

Directed mutagenesis of HI0365. A 4019-bp region of the *H. influenzae* Rd KW20 genome consisting of the CDSs designated HI0364 to HI0367 was amplified by PCR using primers 0364CloneF and 0364CloneR (Table 4.2). The PCR product was gel purified and cloned into pCR2.1-TOPO to yield pTV10. The construct was transformed into DH5α by electroporation and plated on LB agar containing ampicillin, kanamycin, and X-gal. Preliminary verification of the correct insert was performed by digestion of the plasmid at the unique *Eco*RI sites flanking the TA cloning site and size analysis by agarose gel electrophoresis. These results were confirmed by automated DNA sequencing. The *Eco*RV-excised spectinomycin cassette from pSPECR (158) was cloned into the unique *Swa*I site (located within HI0365) of TV10, transformed into chemically competent TOP10 and plated on LB agar containing spectinomycin to yield pTV15. The presence of the correct construct was confirmed by restriction digestion and DNA sequence analysis. The mutant allele was transformed into *H. influenzae* Rd KW20 using the static aerobic transformation method described above. Following overnight growth, transformants were selected on sBHI containing spectinomycin. Chromosomal DNA from selected transformants was prepared, and the presence of the correct construct in the

H. influenzae genome was confirmed by PCR size analysis using primers QPCR-0365-F and QPCR-0366-R2 (Table 4.2).

Directed mutagenesis of HI0939. An insertion into HI0939 was created in the same manner as was used for the mutagenesis of HI0365. A 2614-bp region of the *H. influenzae* Rd KW20 genome consisting of the CDSs designated HI0937 to HI0940 was amplified by PCR using primers 0938CloneF and 0938CloneR (Table 4.2) and cloned into pCR2.1-TOPO to result in pTV05. The spectinomycin cassette from pSPECR was cloned into the unique *Swa*I site in pTV05 (located in HI0939) to create pTV23. Following transformation into Rd KW20 by the MIV method, the presence of the correct insertion in the chromosomal DNA was verified by PCR analysis using primers QPCR-0939-F and 0938Clone-R.

Directed mutagenesis of HI0366. A 1384-bp region immediately upstream of HI0366 was amplified from the Rd KW20 genome by PCR using primers 0366-UP-A and 0366-UP-B and cloned into pCR2.1-TOPO to create pTV16. These primers were designed to incorporate unique *Eco*RI and *Bam*HI sites to facilitate construction of the mutagenic construct. A 934-bp region immediately downstream of HI0366 was amplified by PCR using primers 0366-DN-A and 0366-DN-B and cloned into pCR2.1-TOPO to create pTV17. These primers were designed to include unique *Hin*dIII and *Bam*HI sites. The insertion in pTV16 was excised by restriction digestion using *Eco*RI and *Bam*HI. The downstream region in pTV17 was excised by restriction digestion using *Bam*HI and *Hin*dIII. In addition, pUC18N was digested with *Eco*RI and *Hin*dIII. The digestions were resolved by agarose gel electrophoresis and the appropriate bands were removed, purified using the Geneclean III kit and resuspended in 10 µl HPLC-grade H_2O.

A construct consisting of a vector backbone and the flanking regions of HI0366 was assembled by ligation of equimolar amounts of digested pUC18N and the upstream and downstream fragments and followed by transformation into DH5α by electroporation. An isolate containing a plasmid exhibiting a single band of appropriate size upon agarose gel electrophoresis was chosen and designated as pTV18. The spectinomycin resistance cassette from pSPECR was inserted into pTV18 at the unique *Bam*HI site engineered at the junction of the upstream and downstream flanking regions to yield pTV19. The correct construct was verified by restriction digestion and automated DNA sequence analysis. The construct was transformed into *H. influenzae* Rd KW20 by the static aerobic method and plated on sBHI containing spectinomycin. Resistant colonies were chosen and an Rd KW20 mutant lacking HI0366 was verified by PCR size analysis using primers 0366-UP-A and 0366-DN-B and with primers QPCR-0366-F2 and 0366-R.

Directed mutagenesis of HI1182/1183 (*ligA*). The CDSs designated HI1182 and HI1183 in the Rd KW20 genome form a single CDS, containing a frameshift mutation in the original sequence data, and encodes an ATP-dependent DNA ligase (LigA) (32). Deletion of HI1182/HI1183 was performed as described above for the deletion of HI0366. Briefly, a 629-bp region immediately upstream of the *ligA* gene was amplified from Rd KW20 chromosomal DNA by PCR using primers 1182-UP-A and 1182-UP-B and cloned into pCR2.1-TOPO to create pTV20. A 513-bp region downstream of *ligA* was PCR amplified from the Rd KW20 chromosome using primers 1183-DN-A and 1183-DN-B and cloned into pCR2.1-TOPO to create pTV21. The upstream and downstream regions were cloned into pUC18N to create pTV22. The spectinomycin cassette from pSPECR was cloned into the unique *Bam*HI site engineered at the junction

of the upstream and downstream fragments yielding pTV23. The deletion construct was transformed into *H. influenzae* Rd KW20 using the MIV technique and plated on sBHI containing spectinomycin. Resistant colonies were chosen and an *H. influenzae* mutant lacking *ligA* was verified by PCR size analysis using primers 1182-UP-A and 1183-DN-B and with primers 1182-F and 1183-R. Primers 1182-F and 1183R were also used to PCR amplify the entire *ligA* gene for cloning into pCR2.1-TOPO. The resulting plasmid, pTV31, was sequenced to determine the correct sequence of *ligA*.

DNA binding and uptake analysis. Radiolabeled DNA for use in assays to test the ability to bind and uptake transforming DNA was prepared by nick translation as described by Dougherty and Smith (42). In order to digest the 3' termini of the DNA, 12 μg of MAP9 DNA was incubated at 37°C for 25 minutes in a 100 μl reaction containing 15 units T4 DNA polymerase (Fisher Bioreagents, Fairlawn, N.J.), 1X NEB Buffer 4 [20mM Tris-acetate, 10 mM magnesium-acetate, 50 mM potassium acetate, 1 mM DTT (pH 7.9)] and 0.1 mg/ml BSA. This was followed by the addition of dGTP, dCTP and dTTP (100 μm final concentration) and 30 μCi [α-^{32}P]dATP (3000 Ci/mmole), and the reaction was incubated at 37°C for 20 minutes. Unincorporated label was removed from the chromosomal DNA by Sephadex G-50 column chromatography. The volume of labeled DNA mixture was adjusted to 500 μl with TE buffer (specific activity 1.5 x 10^5 cpm/μg). Mutant and wild-type cultures (5 ml) were made competent by the MIV method as described above. Prior to addition of radiolabeled DNA, 1 ml was removed to determine transformation frequency. Ten microliters of radiolabeled DNA (250 μg) was added to 1 ml of competent cells and the mixture was incubated for 10 minutes at 37°C with shaking. The samples were transferred to an ice bath and divided

equally into two tubes to test DNA binding and DNA uptake. The sample for DNA binding was centrifuged and washed once with 1 ml MIV media and resuspended in 100 µl of MIV. The sample for DNA uptake was treated with 10 µg of DNase I for 5 minutes, followed by the addition of NaCl to a final concentration of 0.5 M. The cells were pelleted, washed once in MIV containing 0.5M NaCl and resuspended in 100 µl MIV. The total cell associated count (DNA binding proficiency) and the DNase I-resistant count (DNA uptake proficiency) was quantified on a Beckman LS 6000SC scintillation counter.

Determination of transformation frequencies. One ml of competent cells were incubated with 1 µg of MAP9 DNA for 15 minutes at 37°C followed by addition of 1 µg of DNase I and further incubation for 5 minutes. Transformation efficiency was determined using a quantification method described by Jett *et al*. (74). Briefly, 10 µl samples from each dilution were plated in sextuplicate on square, gridded plates containing sBHI or sBHI-novobiocin. The use of novobiocin negates the need for an incubation period for development of antibiotic resistance. The frequency of transformation was determined by dividing the transformant colonies on the antibiotic-supplemented plates by the viable count data from the control plates. The effect of gene disruption on transformation efficiency was determined by dividing the transformation frequency of the mutant by the rate from a concurrently run Rd KW20 control. The benefit of adapting this quantification method to transformation studies is that more samples can be analyzed with far fewer plates required for statistical reliability, and the plating can be performed much faster than with the traditional methods. The single downside is that the lowest transformation frequency that can be observed is 10^{-7}.

Comparison of this method to the standard protocol indicated that the transformation frequencies observed with both methods was equal for studied samples (data not shown).

Analysis of competence gene transcription in transformation mutants. Samples for analyzing competence gene transcription in the insertion or deletion mutants were removed concomitant with assays of DNA uptake and binding. The samples were obtained immediately prior to and 60 minutes following transfer into MIV media. Samples were preserved and processed into cDNA as previously described. The levels of transcripts of 16s rRNA and other transformation-related genes were determined by Q-PCR and compared between wild-type and transformation-defective mutants as described.

Complementation of TMV24. In an attempt to complement the transformation defect in the HI0939 insertion mutant strain (TMV24), two plasmids were constructed that carried all or part of the predicted HI0938 operon. A 1.5-kbp PCR product, encoding the CDSs HI0938 and HI0939 and the associated CRE site, was amplified from Rd KW20 genomic DNA using primers 0938-41F and 0938-39R. This product was cloned into pCR2.1-TOPO as previously described to create pTV25. This plasmid was digested with *Hin*dIII and *Bam*HI and the DNA band corresponding to the chromosomal insert was purified by agarose gel electrophoresis as previously described. The purified band was subcloned into *Hin*dIII-*Bam*HI digested pSU2718, a shuttle vector with a p15a origin of replication that allows establishment of the plasmid in *H. influenzae.* This construct was electroporated into *E. coli* DH5α and a plasmid bearing the correct insertion was recovered and designated pTV27. A 2.4-kbp PCR product, encoding the CDSs HI0938 to HI0941 and the associated CRE site, was amplified from Rd KW20 genomic DNA using

primers 0938-41F and 0938-41R. This product was cloned into pCR2.1-TOPO to create pTV26. The insert from pTV26 was then subcloned into pSU2718 as described above to create pTV28. While the parent vector pSU2718 was successfully used to transform Rd KW20, repeated attempts to establish pTV27 and pTV28 into Rd KW20 and TMV24 were unsuccessful using the both the MIV/glycerol and electroporation methods.

Complementation of TMV19. To complement the transformation defect in the HI0366 deletion mutant strain (TMV19), a plasmid was constructed that carried the CDS HI0365 and HI0366 and the associated CRE site. A 2.1-kbp PCR product was amplified from Rd KW20 genomic DNA using primers 0365-6F and 0365-6R. This product was cloned into pCR2.1-TOPO as previously described to create pTV29. This plasmid was digested with *Hin*dIII and *Bam*HI and the DNA band corresponding to the chromosomal insert was purified by agarose gel electrophoresis as described. The purified band was subcloned into *Hin*dIII-*Bam*HI digested pSU2718 to create pTV30. This plasmid was electroporated into TMV19 to create the strain TMV30. PCR analysis confirmed that the original deletion was maintained in TMV30. Wild-type Rd KW20, TMV19 and TMV30 were compared in their ability to be transformed by MAP9 DNA to novobiocin resistance using the MIV method of competence induction.

Additional analysis of the HI1161 mutant. Disruption of HI1161 by Tn5, recovered in the global transposon mutagenesis strategy described in Chapter Two, resulted in a mutant with lowered transformation frequencies. In order to ensure that the defect was the result of the Tn5 insertion in HI1161, a PCR product containing the transposon and approximately 1200-bp of DNA flanking the insertion site was generated from TMV1874-1 chromosomal DNA using primers 1161CloneF and 1161CloneR. This

fragment was gel purified, cloned and sequenced to verify the presence of the insertion. In addition, the gel purified PCR product was retransformed into Rd KW20 and transformants assayed to confirm the transformation defect. TMV1874-1 was also compared to Tn5 insertion mutants (TMV1778, TMV1767) recovered in the genes located immediately downstream of HI1161.

In order to determine whether lowered transformation frequencies in the HI1161 mutants were due to altered levels of intracellular cAMP, TMV1874-1 was assayed to determine if addition of exogenous cAMP could overcome the transformation defect. Competence was induced in TMV1874-1 cells as previously described except 1 mM cAMP was added to the MIV media (160). The transformation frequency was determined as above.

Nucleotide sequence accession number. The nucleotide sequence of the complete HI1182/HI1183 (*ligA*) gene has been deposited in the GenBank database under accession number AY662955.

RESULTS

Analysis of the Tn5 insertion library. The creation and mapping of transposon insertions in Rd KW20 genomic DNA is described in Chapter Two of this work. The final Tn5 library was composed of plasmids bearing insertions in 245 different genes and intergenic spaces (ignoring rRNA insertions). Analysis of the library indicated that multiple insertion sites were located in some of the genes. These Tn5 insertions were used to mutate Rd KW20 by additive genetic transformation. Kanamycin- or chloramphenicol-resistant colonies were then examined for their ability to be transformed

to novobiocin or streptomycin resistance using the static aerobic method. The methodology utilized for this procedure allowed this screening protocol to be performed in a highly-parallel manner (Figure 4.2). Colonies displaying a transformation-defective phenotype were extensively rescreened using the standard MIV method for competence induction. Mutants with greater than a 3-fold decrease in transformation frequency using the MIV method were selected for further studies. Four Tn5 insertions resulted in mutants with lowered transformation frequencies. An insertion in *fruB* resulted in a mutant with a 4-fold decrease in transformation. The transformation defective phenotype of *fruB* mutants has been previously described (93, 95). Several insertions located in *atpD*, which encodes the β chain of the F_1 subunit of the ATP synthase complex, resulted in a 10-20-fold decrease in transformation frequency. Transformation defects had previously been noted for insertions located in *atpA* and *atpB*, encoding subunits of both the F_0 and F_1 subunits of the ATP synthase complex (61); therefore, the *atpD* defects could be predicted. One transformation-defective Tn5 insertion mutant was located in a gene not previously implicated in transformation in any organism and resulted in an approximately 25-fold decrease in transformation frequency. The predicted product of this gene (HI1161) is a highly conserved protein with no known function. Studies of this mutant are described in greater detail later in this work.

Computer analysis of CRE regions. Previous analysis of the Rd KW20 genomic sequence identified putative CRE sites upstream of the CDSs HI0364 and HI1181 (80). An assumption was made that these CREs were associated with HI0364 and HI1181 although the rationale for these assignments is unclear. However, both CRE sites are located between divergently transcribed operons and could theoretically effect

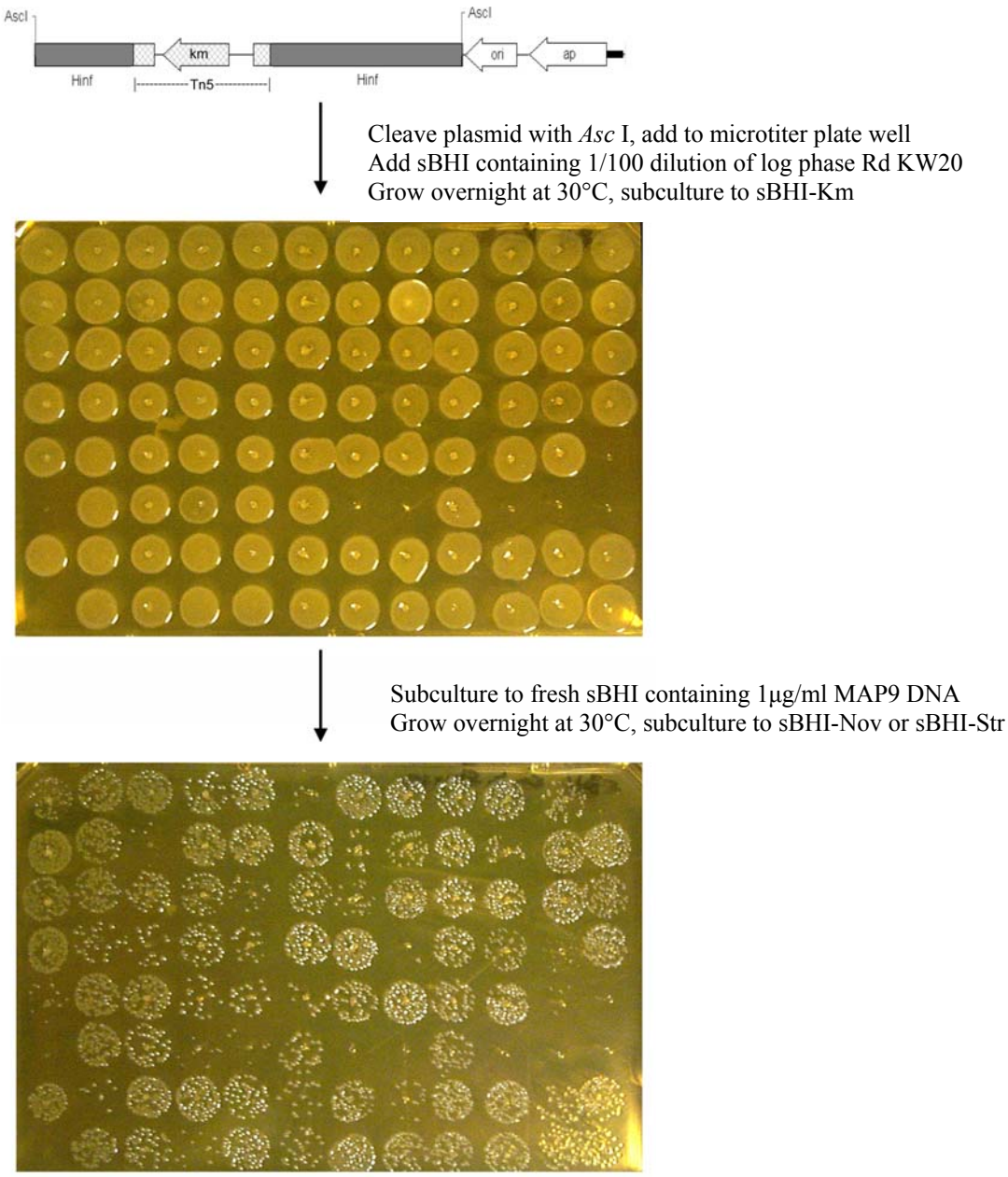


Figure 4.2 – Screening transposon libraries for transformation mutants. Chromosomal inserts were excised from the plasmid backbone and transformed into Rd KW20 using the static aerobic method. Transformants were selected by overnight growth on sBHI-Km. Kanamycin-resistant colonies were subcultured to fresh sBHI containing MAP9 DNA and grown overnight at 30°C. Transformants were detected by subculturing onto sBHI-Nov and sBHI-Str. Mutants that demonstrated lower numbers of colonies on sBHI-Nov or sBHI-Str were subsequently reexamined to confirm the transformation defect.

transformation in either direction. Additionally, MacFadyen identified a site upstream of HI0937 (*suhB*) in strain Rd KW20 that was categorized as a binding site for the cAMP-CRP complex (94). However, this latter site displays an 85% match with the consensus CRE (Figure 4.1). Blastp analysis of the CDSs surrounding each of these sites indicated that the assigned targets for their control might be incorrect. The product of HI0366 showed significant homology to conserved hypothetical proteins annotated as putative fimbrial biogenesis and twitching motility proteins, suggesting that the protein is pilin-related and therefore a possible competence factor. The gene upstream of HI0366 (HI0365) is a putative Fe-S cluster redox enzyme. The designated subject for CRE0364 control, HI0364, encodes a putative penicillin-binding protein. While insertions in other PBPs have been shown to have a negative effect on transformation in *H. influenzae* (42), presumably due to changes in cell wall structure, none of these genes have been demonstrated to be under the control of the competence regulon and none are associated with apparent CRE elements.

Examination of the CDSs that could be divergently transcribed from the putative CRE0937 also identified potential pilin-related genes. The products of the CDSs designated HI0938 to HI0941 share little homology to proteins outside of members of the Pasteurellaceae. However, characteristics of these four putative proteins are consistent with prepilin proteins: their mass is less than 20kDa, they contain a short N-terminal leader peptide and a hydrophobic stretch, and they display no sequence conservation in the C-terminal regions (28). In contrast, HI0937 shows significant homology (65% identity) to *suhB* from *E. coli,* a gene that encodes an inositol monophosphatase that also appears to participate in posttranscriptional control of gene expression (31).

Examination of the CDSs surrounding the putative CRE upstream of HI1181 indicated that HI1182/HI1183 is a more likely candidate to fall under the control of the competence regulon. HI1182/HI1183 encodes an ATP-dependent DNA ligase (LigA), whereas HI1181 encodes a phosphoheptose isomerase involved in lipooligosaccharide biosynthesis (22, 32).

While natural genetic transformation has not been demonstrated in all the members of the Pasteurellaceae, limited examination of the partial or complete genomic sequences available for members of the family identified homologues of *rec-2*, *dprA*, *comM*, and the *com* and *pil* operons (data not shown). Furthermore, putative CRE sites could be located upstream of all of these genes. To gain further insight into the organization and potential regulon involvement of the CDSs contiguous with the three putative CRE sites discussed above, the homologous regions in the genomic sequences of the other members of the family Pasteurellaceae (Table 4.3) were located and analyzed. The sequence of the HI0365 and HI0366 homologs in *M. haemolytica* is incomplete and could not be analyzed. HI0365 and HI0366 homologs are present in the sequences from the other analyzed Pasteurellaceae and are associated with a putative CRE site immediately upstream (Figure 4.3). In *H. somnus, H. ducreyi*, and *M. haemolytica* the CDSs downstream of HI0366 in Rd KW20 are located elsewhere in the genome and do not have an associated CRE site. Only *A. actinomycetemcomitans* and *M. haemolytica* have HI0364 homologs, and these homologs are not associated with an identifiable CRE in either organism. The putative prepilin homologs HI0938 – HI0941 and *recC* were found together, with an associated CRE site, in all the examined genomic sequences (Figure 4.4). In contrast, the homolog of HI0937 in each of the organisms is located

elsewhere in the genome and is not associated with an identifiable CRE in any of the family members. Besides Rd KW20, only *A. actinomycetemcomitans* and the *H. somnus* strains contain *ligA* (HI1182/HI1183) homologs; in these organisms a putative CRE site is located immediately upstream of the gene (Figure 4.5). HI1181 homologs exist in each of the family members but are not associated with an identifiable CRE site.

TABLE 4.3 – Pasteurellaceae genomic sequences examined in this study

Organism	Accession or Progress	Website
Completed:		
Haemophilus influenzae Rd KW20	NC_000907	http://www.tigr.org
Haemophilus ducreyi 35000HP	NC_002940	http://www.microbial-pathogenesis.org
Pasteurella multocida PM70	NC_002663	http://www.cbc.umn.edu/
In progress:		
Actinobacillus actinomycetemcomitans HK1651	Assembled	http://www.genome.ou.edu
Actinobacillus pleuropneumoniae serotype 1 str. 4074	8.6 x	http://www.micro-gen.ouhsc.edu
Haemophilus somnus 129PT	10.7x	http://www.jgi.doe.gov
Haemophilus somnus 2336	8.2 x	http://www.micro-gen.ouhsc.edu
Mannheimia haemolytica PHL213	6 x	http://www.hgsc.bcm.tmc.edu

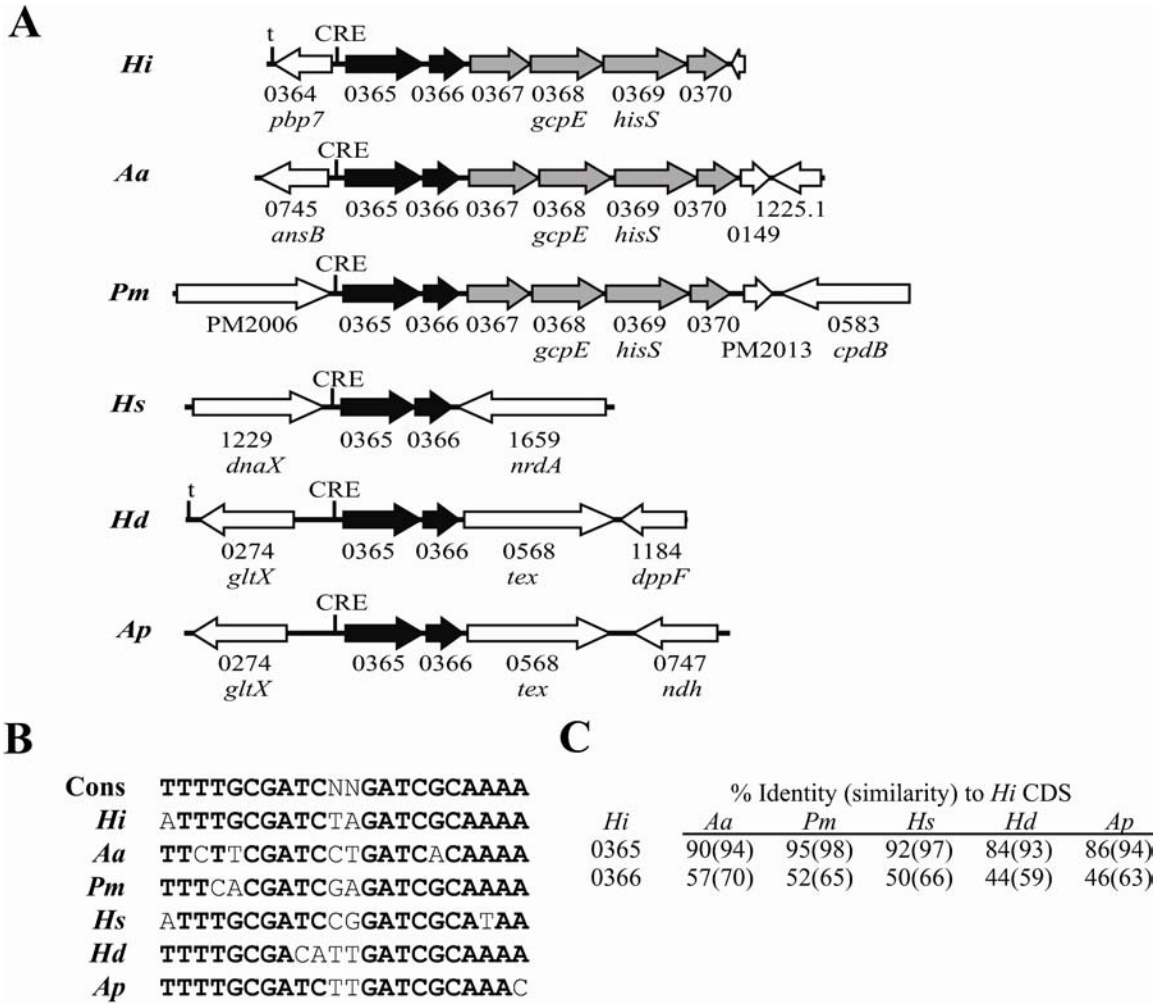
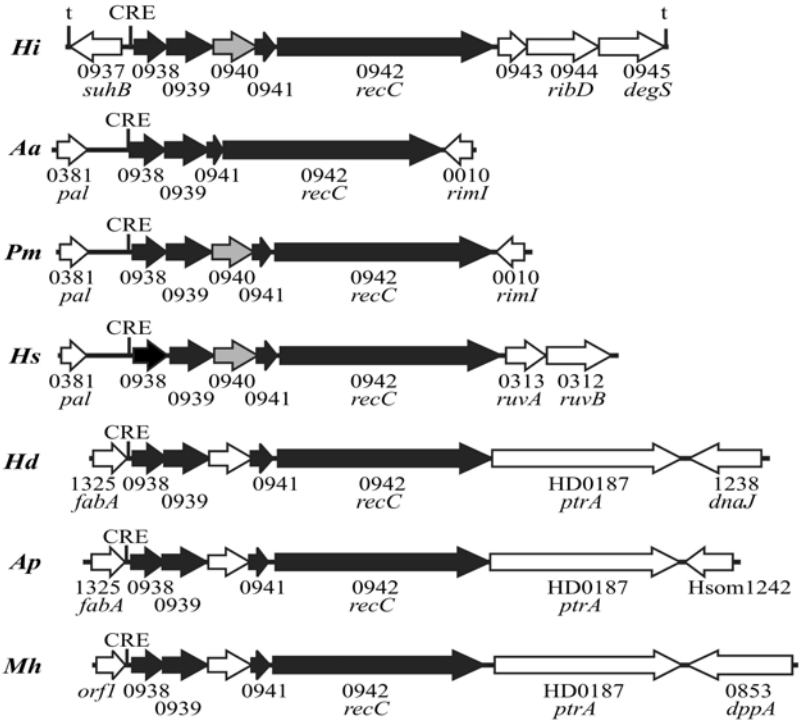


Figure 4.3 – Organization and conservation of the CRE0364 region in the family Pasteurellaceae.

A. Numbers shown correspond to the HI number in the Rd KW20 annotation unless otherwise indicated and are based upon the closest homolog in the Rd KW20 sequence. Solid colors indicate genes whose location in reference to the CRE element is conserved across the family. Shaded colors indicate genes with partial positional conservation **B.** Comparison of the putative CRE sequences upstream of the HI0365 homolog in each organism compared to the consensus CRE sequence from Rd KW20 (Cons). Nucleotides in bold share 100% identity with the consensus sequence. **C.** Degree of identity of proteins homologous to HI0365 and HI0366 in the Pasteurellaceae. Abbreviations used: *Hi*: *H. influenzae* Rd KW20; *Aa*: *A. actinomycetemcomitans* HK1651; *Pm*: *P. multocida* PM70; *Hs*: *H. somnus* strains 129PT and 2336; *Hd*: *H. ducreyi* 35000HP; *Ap*: *A. pleuropneumoniae* serotype 1 str. 4074

A



B

Cons	**TTTTGCGATCNNGATCGCAAAA**
Hi	C**TTTGCGAT**ACA**GATCGC**K**AAA**
Aa	C**TTTGCGATC**CT**G**C**TCGCAAAA**
Pm	**TTTTGCGATC**TGC**ATC**T**CAAAA**
Hs	**TTTTGCGATC**TGT**ATCG**T**AA**TT
Hd	**TTTTGCGATC**AA**GATCG**TG**AAA**
Ap	**TTTTGCGATC**AA**GATCG**A**A**T**AA**
Mh	**TTTTGCGATC**TGC**ATCG**A**AAAA**

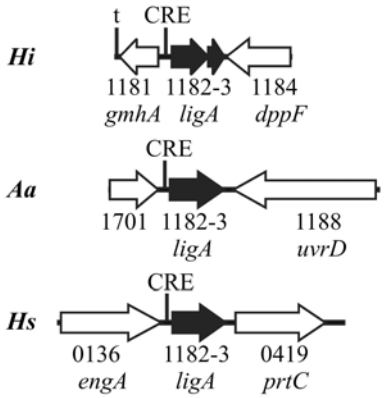
C

% Identity (similarity) to *Hi* CDS

Hi	*Aa*	*Pm*	*Hs*	*Hd*	*Ap*	*Mh*
0938†	59(74)†	42(63)†	53(76)†	29(44)†	27(43)†	27(45)†
0939‡	52(72)‡	47(66)‡	51(71)‡	29(48)‡	29(49)‡	26(47)‡
0940	Np	36(53)	43(61)	Np	Np	Np
0941	56(71)	46(67)	46(62)	30(52)	32(52)	34(60)
recC	52(68)	48(65)	46(64)	39(57)	40(59)	38(56)

Figure 4.4 – Organization and conservation of the CRE0937 region in the family Pasteurellaceae. **A.** Numbers shown correspond to the HI number in the Rd KW20 annotation unless otherwise indicated and are based upon the closest homolog in the Rd KW20 sequence. Solid colors indicate genes whose location in reference to the CRE element is conserved across the family. Shaded colors indicate genes with partial positional conservation. **B.** Comparison of the putative CRE sequences upstream of HI0938 in each organism compared to the consensus CRE sequence from Rd KW20 (Cons). Nucleotides in bold share 100% identity with the consensus sequence. **C.** Degree of identity of proteins homologous to HI0938-HI0942 in the Pasteurellaceae. † and ‡ rpsblast database indicates presence of PulG and PulJ domains respectively. Np indicates homolog not present. See Figure 4.3 for abbreviations used, except *Mh*: *M. haemolytica* PHL213.

A



B

Cons	**TTTTGCGATC**NN**GATCGCAAAA**
Hi	**TTTTGCGAT**TTA**GATCG**A**AAAA**
Aa	A**TTTG**G**GATC**GCCGA**CGCAAAA**
Hs	**TTTTGCGATC**TCTG**TCGCAAA**T

Figure 4.5 – Organization and conservation of the CRE1181 region in the family Pasteurellaceae. - A. Numbers shown correspond to the HI number in the Rd KW20 annotation unless otherwise indicated and are based upon the closest homolog in the Rd KW20 sequence. Solid colors indicate genes whose location in reference to the CRE element is conserved across the family. The *A. actinomycetemcomitans* and *H. somnus* LigA homologs demonstrated 69% identity (89% similarity) and 71% identity (85% similarity), respectively, to the Rd KW20 LigA sequence. **B.** Comparison of the putative CRE sequences upstream of the *ligA* homolog in each organism compared to the consensus CRE sequence from Rd KW20 (Cons). Nucleotides in bold share 100% identity with the consensus sequence. See Figure 4.3 for definitions of the abbreviations used.

Q-PCR examination of gene expression during competence development. To determine if the pattern of putative CRE relationships identified *in silico* was biologically relevant, the expression of genes during the development of competence in MIV media was investigated. Previous studies using β-galactosidase fusions demonstrated increased production of the fusion products from the CRE sites upstream of *comA* and *rec-2* in response to competence development (61). Expression from *tfoX* fusions also increased during competence development (169). However, Bannister demonstrated that this effect appears to be mediated primarily by RNA secondary structure in the *tfoX* transcript (14). Examination of the expression of *comA* and *rec-2* by Q-PCR demonstrated that the expression of the two genes increased dramatically following transfer into MIV media (Figure 4.6). Maximal increases of 117.1-fold (range 104.3-131.5) for *rec-2* and 97.5-fold (range 86.2-110.2) for *comA* were observed 60 minutes into competence induction. Subsequently, expression of both genes declined to pre-treatment levels at 160 minutes. In contrast, expression of *tfoX* remained relatively constant, with a maximum decrease of 2.6-fold (range 2.1-3.3) from pre-treatment levels.

Examination of genes contiguous with CRE0364 (Figure 4.7) showed that transcription of the three genes studied appeared to drop immediately upon centrifugation but recovered to near pretreatment levels at 20 minutes. Transcription of HI0364 (*pbp-7*) and HI0368 (*gcpE*) did not increase more than 1.1-fold (range 0.88-1.3) and 1.4-fold (range 1.3-1.6) over the control levels. However, expression of HI0366 reached a maximal increase of 5.3-fold (range 4.8-5.9) at 60 minutes after transfer into MIV. The transcription levels of all three examined genes decline below the pretreatment levels after 120 minutes in MIV. Transcription of HI0365 was also examined and it

demonstrated the same magnitude of increase as HI0366 at 60 minutes but transcript levels after 120 minutes fell below the linear range determined for the primers utilized (data not shown).

Examination of genes contiguous with CRE0937 (Figure 4.8) showed a transcription profile similar to *rec-2* and *comA* for HI0938 and HI0939. Maximal expression was observed at 60 minutes into induction with increases of 82.5-fold (range 75.0-90.7) and 35.6-fold (range 29.0-43.7), respectively. In contrast, expression of HI0937 (*suhB*) and HI0942 (*recC*) remained relatively steady-state and showed maximal increases of 2.1-fold (range 1.8-2.6) and 1.8-fold (range 1.3-2.5), respectively, following transfer into MIV.

Finally, expression of HI1181 (*gmhA*) and HI1182 (*ligA*), transcribed divergently from CRE1181 was examined (Figure 4.9). Transcription of *ligA* demonstrated a pattern similar to *rec-2* and *comA* and had a maximal 61.1-fold (range 52.3-71.4) increase at 60 minutes after transfer into MIV. Transcription of *gmhA* decreased 3.3-fold (range 2.7-4.4) upon transfer into MIV and remained below the pre-treatment level in the remainder of the samples.

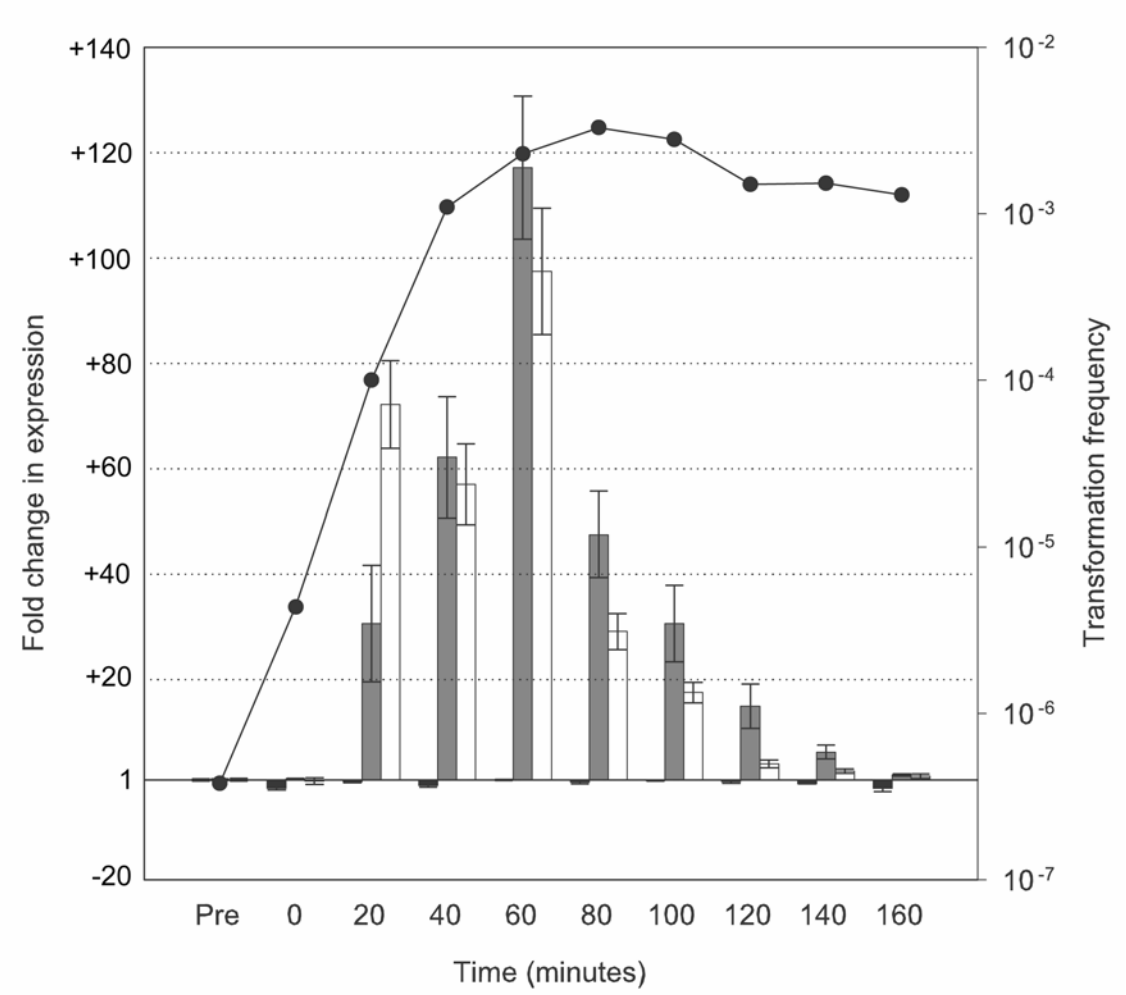


Figure 4.6 – Expression profile of *tfoX, rec-2* and *comA* during competence development. Q-PCR was employed to examine transcription of known competence genes during development of competence in MIV media. Values represent the fold change in expression compared to levels in the control (pre-treated) sample as determined by the $2^{-\Delta\Delta Ct}$ method (89). Error bars indicate the range in fold change as determined by this method. *tfoX* is represented by black bars, *rec-2* by gray bars and *comA* by white bars. CRE sites are present upstream of *rec-2 and comA* and both have been shown to be dependent on *tfoX* for expression. The transformation frequency at each timepoint is shown by the closed circles.

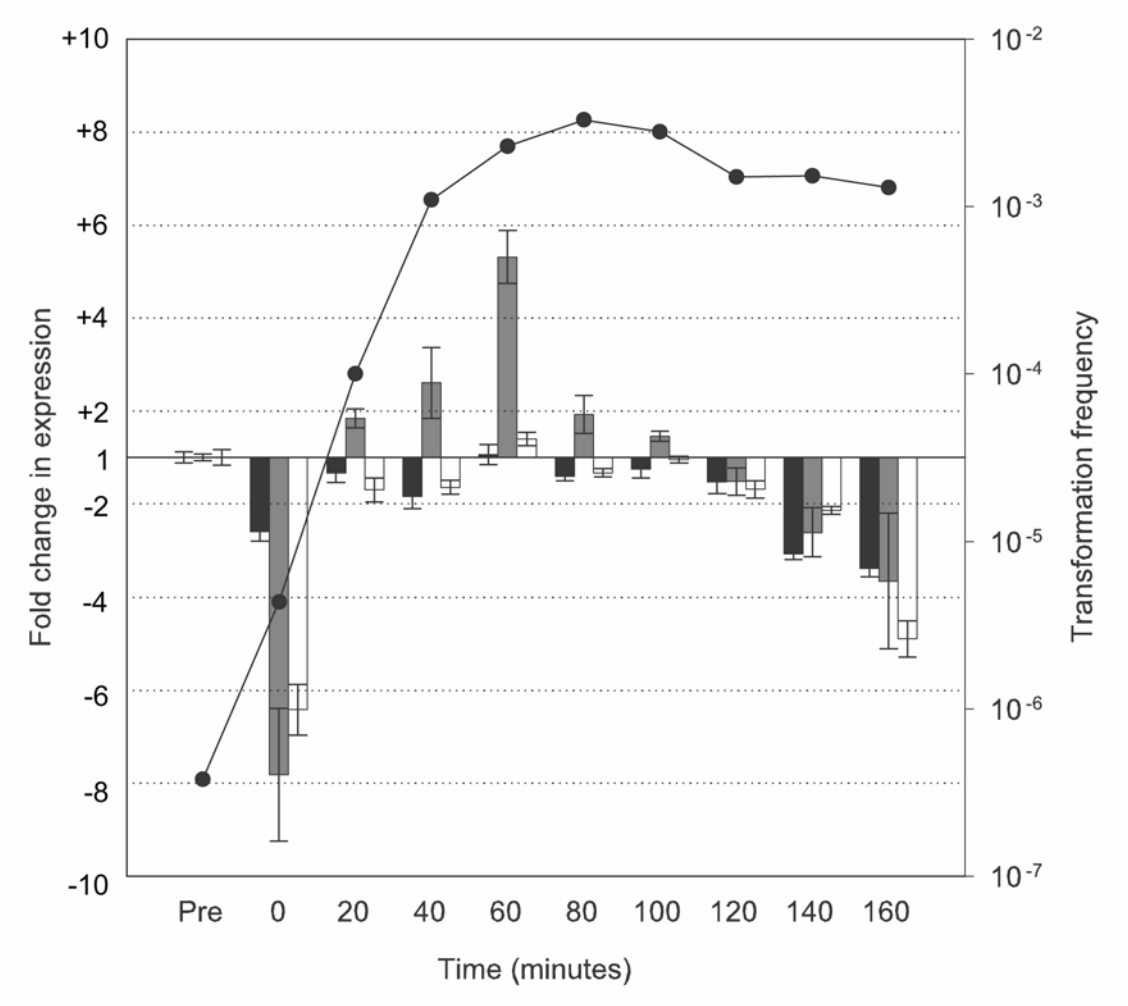


Figure 4.7 – Expression profiles of genes contiguous with CRE0364 during competence development. Q-PCR was employed to examine transcription of HI0364 (black), HI0366 (gray), and HI0368 (white) during competence development in MIV media. Values represent the fold change in expression compared to levels in the control (pre-treated) sample as determined by the $2^{-\Delta\Delta Ct}$ method (89). Error bars indicate the range in fold change as determined by this method. The transformation frequency at each timepoint is shown by the closed circles.

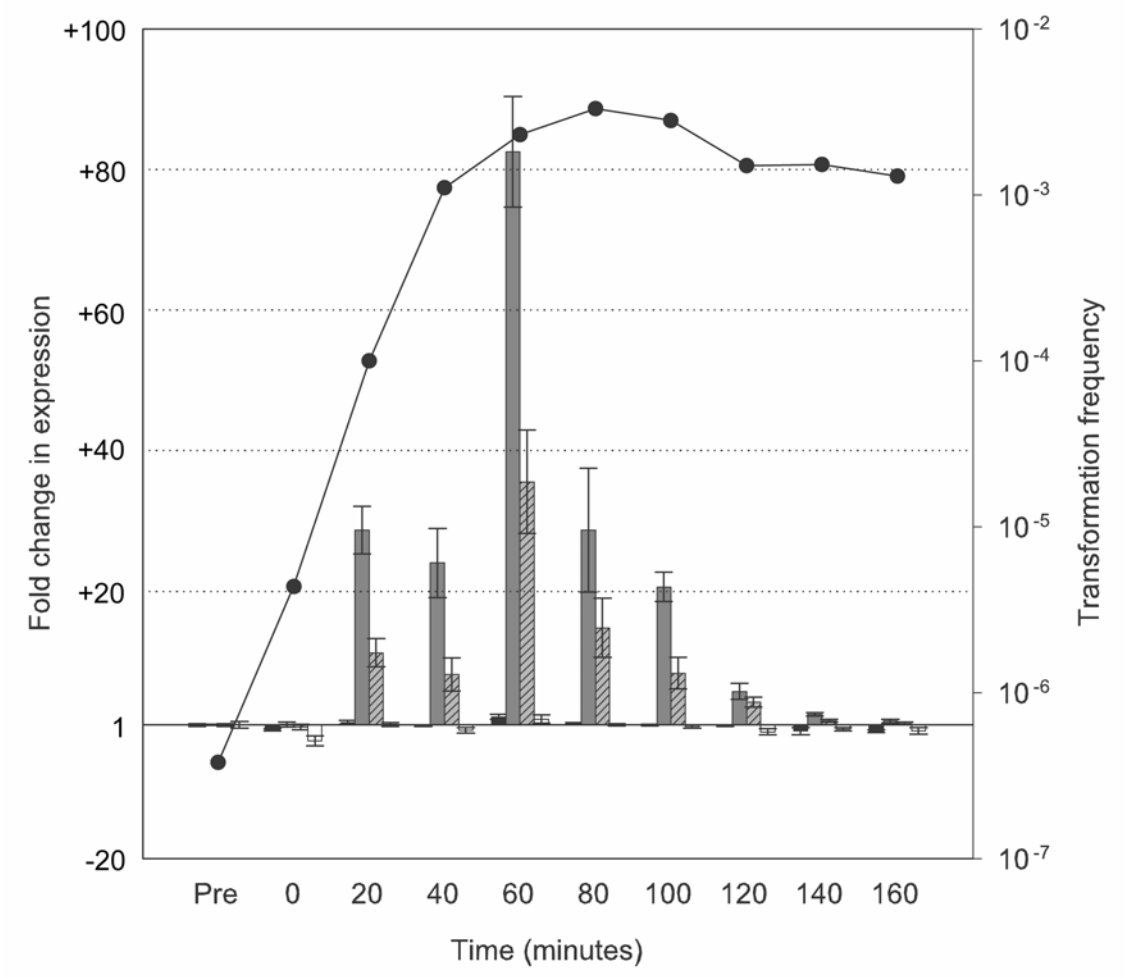


Figure 4.8 - Expression profiles of genes contiguous with CRE0937 during competence development. Q-PCR was employed to examine transcription of HI0937 (black), HI0938 (gray), HI0939 (hatched gray), and HI0942 (white) during competence development in MIV media. Values represent the fold change in expression compared to levels in the control (pre-treated) sample as determined by the $2^{-\Delta\Delta Ct}$ method (89). Error bars indicate the range in fold change as determined by this method. The transformation frequency at each timepoint is shown by the closed circles.

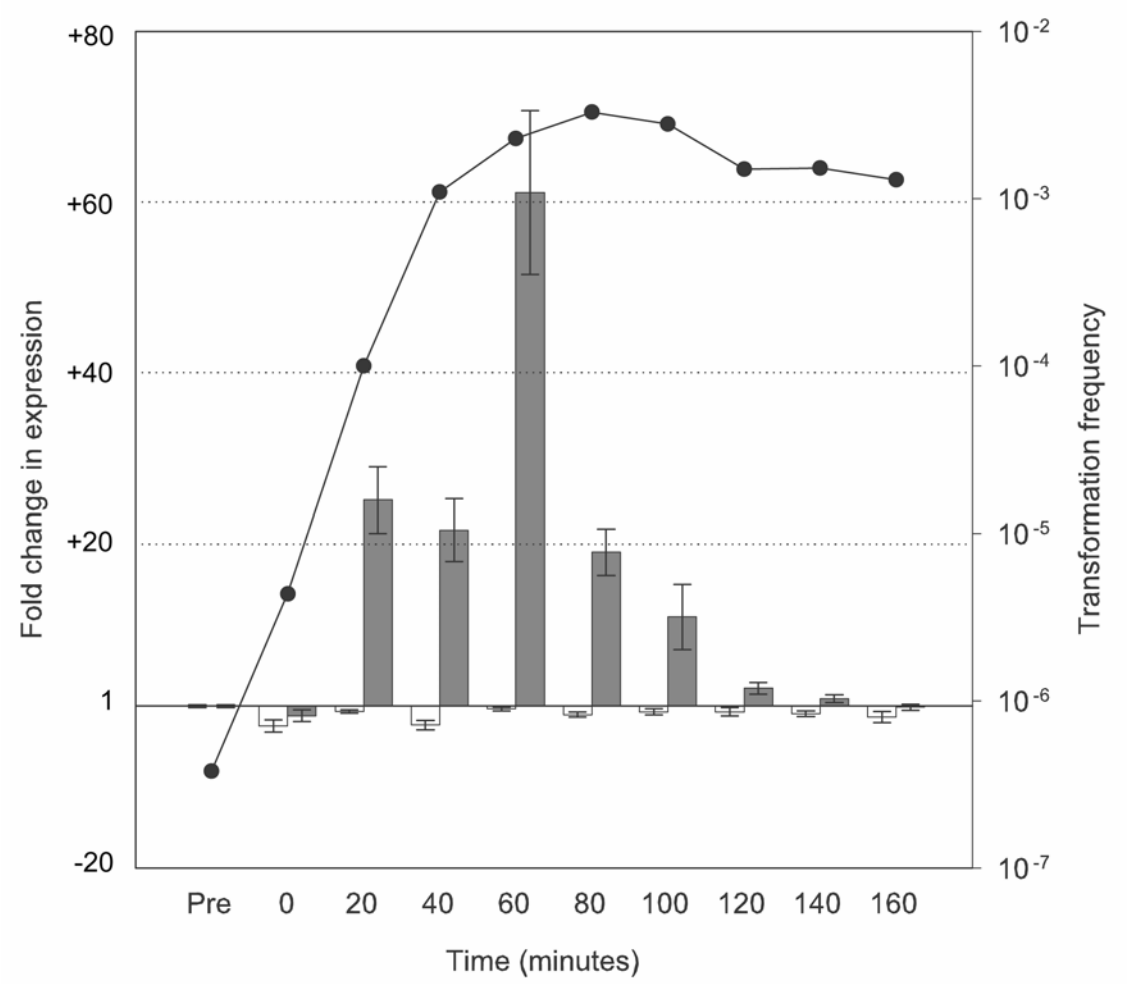


Figure 4.9 – Expression profiles of genes contiguous with CRE1181 during competence development. Q-PCR was employed to examine transcription of HI1181 (white) and HI1182 (gray) during competence development in MIV media. Values represent the fold change in expression compared to levels in the control (pre-treated) sample as determined by the $2^{-\Delta\Delta Ct}$ method (89). Error bars indicate the range in fold change as determined by this method. The transformation frequency at each timepoint is shown by the closed circles.

Characterization of transformation phenotypes. The analysis of MIV-dependent transcription of genes contiguous to CRE0364, CRE0937 and CRE1181 indicated that several new genes may be members of the competence regulon. In order to determine the involvement of these genes in transformation, mutants were created in HI0365, HI0366, HI0939 and *ligA*. Transformation frequencies in MIV media and DNA binding and uptake abilities of these mutant strains were compared to the wild-type strain (Table 4.4). The HI0365 mutant strain TMV15 was as proficient as the wild-type in both transformation to novobiocin resistance and in uptake and binding of labeled MAP9 DNA. In contrast, the HI0366 mutant strain TMV19 was severely impaired in the ability to transform ($< 10^{-7}$ transformants/ml) and to uptake DNA (0.3% of wild-type levels) but demonstrated only a moderate decrease in DNA binding (28 % of the wild-type). The HI0939 mutant (strain TMV24) was also deficient in transformation ($< 10^{-7}$ transformants/ml), binding (1.1%) and uptake (0.08%). The *ligA* mutant (strain TMV23) demonstrated only a 5-fold decrease in transformation frequency and near wild-type levels of binding and uptake (84 % and 92 %, respectively).

The levels of transcription of competence related genes were assayed and compared between wild-type Rd KW20 and the isogenic HI0366, HI0939 and *ligA* mutants (Table 4.5). After 60 minutes incubation in MIV media, levels of *rec-2, comA,* and HI0938 were relatively similar between the wild-type and HI0939 and HI1182 mutant strains. Transcription of these three competence genes was notably lower in the HI0366 mutant strain but the level of *tfoX* transcripts was approximately equal to the wild-type.

TABLE 4.4 – Examination of transformation efficiency, DNA binding and uptake for wild-type and mutant strains of *H. influenzae* Rd KW20

Strain	Genotype	Transformation frequency[a]	DNA binding[b]	DNA uptake[b]
Rd KW20-MIV	wild-type	1.0×10^{-3}	100	100
Rd KW20-BHI	wild-type	$< 1.0 \times 10^{-7}$	0.0	0.0
Rec-1	*rec-1*	$< 1.0 \times 10^{-7}$	102.3	91.9
TMV15	HI0365::spec	1.0×10^{-3}	107.2	110.8
TMV19	ΔHI0366	$< 1.0 \times 10^{-7}$	28.3	0.3
TMV23	HI0939::spec	$< 1.0 \times 10^{-7}$	1.1	0.08
TMV24	ΔHI1182	2.0×10^{-4}	84.2	94.0
TMV1874-1	HI1161::Tn5	4.2×10^{-5}	24.4	4.8
TMV1874-2	HI1161::Tn5	3.3×10^{-5}	14.2	1.8

[a] Number of novobiocin resistant CFU divided by total CFU
[b] Cell associated cpm for strain divided by cell associated cpm for Rd KW20-MIV

TABLE 4.5 – Examination of gene transcription in wild-type and mutant strains of H. influenzae Rd KW20 after 60 minutes incubation in MIV media

		Gene			
Strain	Genotype	*rec-2*	*comA*	HI0938	*tfoX*
Rd KW20	wild-type	1.0 (0.90-1.1)	1.0 (0.91-1.1)	1.0 (0.88-1.1)	1.0 (0.68-1.5)
TMV19	ΔHI0366	*5.1 (4.0-6.4)*	*10.6 (8.2-13.8)*	*5.4 (4.9-6.1)*	1.4 (0.86-2.2)
TMV24	HI0939::Sp	*2.4 (1.7-3.3)*	*1.1 (0.59-2.2)*	1.2 (1.0-1.4)	nt
TMV23	ΔHI1182	*2.5 (2.0-3.3)*	1.4 (1.0-2.0)	1.1 (0.94-1.4)	nt
TMV1874-1	HI1161::Tn5	*4.3 (2.9-6.3)*	*4.8 (3.6-6.5)*	*7.6 (5.8-9.9)*	*2.4 (2.1-2.7)*
TMV1874-2	HI1161::Tn5	*10.5 (8.4-13.1)*	*13.0 (11.1-15.1)*	*18.0 (12.5-25.9)*	1.0 (0.74-1.4)

Data is presented as fold-change in transcription compared to Rd KW20 as determined by the $2^{-\Delta\Delta Ct}$ method. Numbers in parentheses indicate the range of values. Numbers in black and red (*italicized*) indicate increased or decreased transcription, respectively, compared to wild-type. nt= not tested

The lack of a discernable transformation defect in the mutant harboring a spectinomycin cassette in HI0365 is interesting considering that the insertion would likely have produced polar effects on the transcription of HI0366. An examination of the cassette nucleotide sequence failed to locate any potential rho-independent transcriptional terminators. One possible explanation for this phenomenon is that transcription read-through from the cassette was sufficient to allow expression of HI0366. In order to examine this hypothesis, RNA was isolated from Rd KW20 and HI0365::Spec cultures immediately prior to and 60 minutes following transfer into MIV media. Q-PCR examination of HI0366 transcription indicated that expression of this gene was 13.9-fold higher (range 9.9-19.6) in the mutant strain than in Rd KW20 prior to MIV treatment (Table 4.6). Expression of *comA* was also higher in the mutant at this time [3.0-fold (range 1.4-6.2)] but this increase may not be significant as similar increases are also present in the other three mutants (data not shown). At 60 minutes into competence induction, expression of HI0366 and *comA* were roughly equivalent in both the wild-type and mutant. Thus, it appears that insertion of the spectinomycin cassette into HI0365 did not cause the anticipated polar effects on the downstream genes.

TABLE 4.6 – Examination of *comA* and HI0366 transcription in Rd KW20 and TMV15

			Gene	
Strain	Genotype	Sample	*comA*	HI0366
Rd KW20	wild-type	Pre-MIV	1.0 (0.89-1.1)	1.0 (0.78-1.32)
		60 min	1.0 (0.98-1.0)	1.0 (0.87-1.2)
TMV15	HI0365::Sp	Pre-MIV	3.0 (1.4-6.2)	13.9 (9.9-19.6)
		60 min	*1.3 (0.99-1.8)*	*1.1 (0.87-1.3)*

Data is presented as fold-change in transcription compared to Rd KW20 at the same timepoint as determined by the $2^{-\Delta\Delta Ct}$ method. Numbers in parentheses indicate the range of values. Numbers in black or red (*italicized*) indicate increased or decreased transcription, respectively compared to wild-type.

Complementation of the HI0366 mutant strain TMV19. A plasmid (pTV30) was constructed containing the CDSs HI0365, HI0366 and their associated CRE site in the *H. influenzae* shuttle vector pSU2718. The plasmid was established in the HI0366 deletion mutant TMV19 by electroporation. In order to assess whether the plasmid bearing HI0366 was able to complement the chromosomal deletion of the gene, Rd KW20, TMV19 (ΔHI0366) and TMV30 (ΔHI0366; pTV30) were subjected to the MIV method of competence induction and compared in their abilities to be transformed to novobiocin resistance by MAP9 DNA. The establishment of pTV30 in the HI0366 mutant strain was able to fully complement the deletion and restore transformation to wild-type levels thus confirming that the transformation defect in TMV19 was the result of the HI0366 deletion.

Characterization of the HI1161 mutant. The mutant strain TMV1874-1 was the result of transforming Rd KW20 with pASC1874. This Tn5 insertion was located 61aa from the putative N-terminus of the predicted product of the gene. The strain TMV1874-2 was constructed by transforming Rd KW20 with a 3.8-kbp purified PCR product amplified from TMV1874-1 chromosomal DNA that included the Tn5 insertion and flanking sequences. Both strains demonstrated a 25- to 30-fold lower frequency of transformation to novobiocin resistance after incubation in MIV media and reduced levels of DNA binding and uptake (Table 4.4). Examination of transcriptional profiles of *rec-2, comA* and HI0938 (Table 4.5) demonstrate that transcription of the competence regulated genes is decreased in the HI1161 mutants when compared to wild-type cells. This decrease is not the result of reduced transcription of the putative competence regulator *tfoX.* The transformation defect was not alleviated by addition of 1 mM cAMP

to the MIV media, indicating that reduced competence in the HI1161 mutants was not due to changes in intracellular concentrations of cAMP.

Tn5 mutants were also recovered in two genes located downstream of HI1161 (Figure 4.10). The first was located in HI1160 (*hemH*), encoding a ferrochelatase responsible for inserting Fe^{2+} into protoporphyrin IX (PPIX) (128). While neither the HI1160 mutant or the HI1161 mutant was unable to grow in the presence of 10 μM PPIX, the HI1160 mutant was not impaired in its transformation ability. The second insertion was recovered in HI1159m, a CDS located downstream of HI1160 that contains a natural frameshift. When transformed into Rd KW20, this insertion also had no impact on transformation frequency. Taken together, these insertions confirm that the transformation defect in the HI1161 insertion mutant is not due to polar effects on the downstream genes.

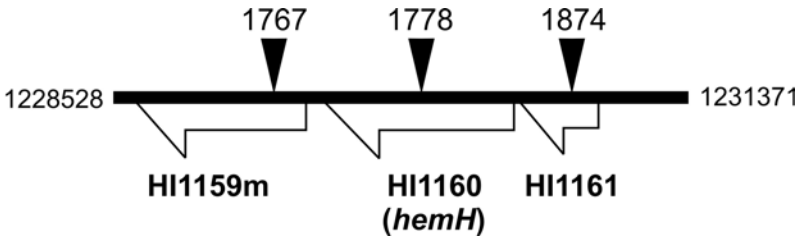


Figure 4.10 – Locations of Tn5 insertions in HI1159m, *hemH,* and HI1161. The above represents the region comprising nucleotides 1228528 to 1231371 in the Rd KW20 genomic sequence. The Tn5 insertion in TMV1767 was located at position 1229407 of the Rd KW20 sequence (codon 17 of 286 in HI1159m). The Tn5 insertion in TMV1778 was located at position 1230015 (codon 164 of 323 in HemH). The Tn5 insertion in TMV1874 was located at position 1230739 (codon 61 of 138 in HI1161). Only the insertion in HI1161 resulted in a transformation-defective phenotype.

DISCUSSION

Three new members of the competence regulon in *H. influenzae* were identified in this study through a combination of *in silico*, transcriptional and mutational analyses. Conserved CRE sites were located upstream of homologs of HI0365, HI0938 and HI1182 in the Pasteurellaceae and transcription of genes within these operons increase in *H. influenzae* Rd KW20 in response to competence induction by transfer into MIV media. Mutational analysis confirmed that the products of HI0366 and HI0939 are involved in transformation in Rd KW20. Additionally, a mutant strain with a Tn5 insertion in HI1161, recovered during a global transposon mutagenesis scheme, had a significant impact on transformation in Rd KW20.

The products of several of the genes found in this study to be upregulated by competence induction exhibit similarities to proteins known to be involved with the type II secreton and type IV pili (Tfp). Together, these are referred to as the PSTC proteins and have been shown to be integral parts of the transformation systems of most bacteria (43). While *H. influenzae* does not produce Tfp, PSTC homologs have been identified in this organism and have a demonstrable role in transformation (42, 149). A model has been suggested by Chen and Dubnau in which binding and uptake is mediated by a pseudopilus structure that utilizes components of the Tfp biogenesis machinery along with competence-specific pilin-like proteins (28).

This work demonstrates that the CRE site originally associated with HI0364 actually controls the transcription of HI0365 and HI0366. Comparative analysis of this operon in other members of the family Pasteurellaceae supports this conclusion. The magnitude of change in transcription of HI0366 was not as large as observed with either

rec-2 or *comA*. This difference might reflect that transcription of this operon is higher in non-competence inducing conditions than other transformation-associated operons since this operon appears to contain a gene unrelated to transformation. Mutational analysis demonstrates that the product of HI0366, but not HI0365, is involved in transformation in Rd KW20. While it might be expected that the insertion in HI0365 would have polar effects on transcription of HI0366, this insertion had no discernable effect on transformation efficiency or DNA binding and uptake. The lack of polar effect is explained by the finding that transcription of HI0366 was sufficient due to read-through from the spectinomycin marker inserted into HI0365. The HI0366 deletion mutant was severely impacted in the ability to transform and to uptake DNA but remained proficient in DNA binding. The phenotype displayed by the HI0366 mutant strain resembles that exhibited by the *com*10 mutant described by Barouki and Smith (18). Since the defective gene in that mutant has not been identified, it is plausible that HI0366 may be the affected gene. The predicted product of HI0366 shares weak homology to PilF from *Pseudomonas aeruginosa* PA01 (29% identity). Mutations in *pilF* in *P. aeruginosa* result in the lack of discernable pili, abolition of twitching motility and the accumulation of processed PilA in the membrane fraction (156). Thus, PilF may be necessary for export or assembly of pilin subunits. Since natural transformation has not been observed in this organism, it is impossible to assess a transformation phenotype in the *P. aeruginosa pilF* mutant. Interestingly, the *P. aeruginosa pilF* locus is nearly identical to that of the HI0366 locus in Rd KW20 (Figure 4.11). No intergenic space is present between *pilF* and the PA01 homolog of HI0367 as is present in the Rd KW20 sequence. However, like the HI0366 mutant, complementation of *pilF* alone is able to restore Tfp production in *P.*

aeruginosa (156). While the similarity of the product of HI0366 to PilF would suggest a similar role in *H. influenzae* as seen in *P. aeruginosa*, the phenotype displayed by the HI0366 mutant is not consistent with this hypothesis. Mutations in other PSTC homologs in *H. influenzae* abolish both binding and uptake of DNA (42, 149) suggesting that an intact pseudopilus structure is necessary for both binding and uptake. Since the HI0366 mutant is proficient in binding DNA, it would suggest a direct role of the protein in transferring DNA across the outer membrane; however, its exact function cannot be identified at this time. Q-PCR examination of expression of several transformation-related genes indicated that the regulation of competence was negatively impacted in the HI0366 mutant. Expression of transformation genes increased after transfer to MIV media but not to the levels observed in the wild-type. The decrease in transcription in the HI0366 mutant was similar to that observed in the HI1161 mutants, yet the impact to DNA uptake and to overall transformation frequencies was considerably different. While DNA uptake was significantly impaired by the insertion in HI1161, DNA was still taken up into a DNase-resistant form and transformation frequency was only reduced 30-fold from the wild-type. In contrast, DNA uptake in the HI0366 mutant was reduced 10-fold greater than that observed in the HI1161 mutants. Additionally, transformation to antibiotic resistance was not observed in the HI0366 mutant, a decrease of greater than four orders of magnitude from wild-type cells. I propose naming HI0366 *pilF2,* and its predicted product PilF2, to recognize the similarity to *P. aeruginosa* PilF but to avoid confusion with the unrelated pilin biogenesis protein PilF of *N. gonorrhoeae.*

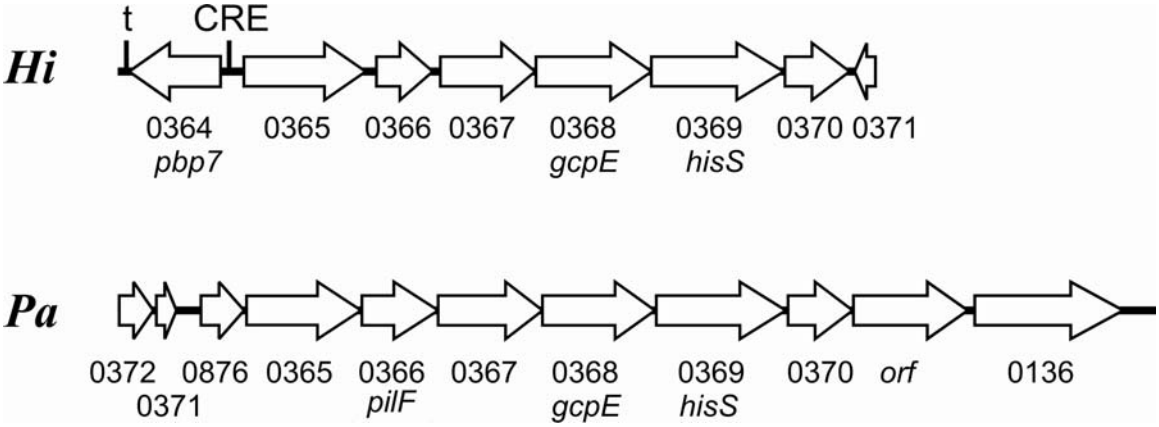


Figure 4.11 – Comparison of the HI0366 locus in *H. influenzae* Rd KW20 with the *pilF* locus of *P. aeruginosa* PA01. Numbers shown correspond to the HI number in the Rd KW20 annotation unless otherwise indicated and are based upon the closest homolog in the Rd KW20 sequence. Abbreviations used: *Hi: H. influenzae* Rd KW20; *Pa: P. aeruginosa* PA01.

This study also supports the inclusion of HI0938 and HI0939 as members of the competence regulon in *H. influenzae*. Although transcription analysis of the CDSs HI0940 and HI0941 was not performed, the similarity of their predicted products to prepilin-like proteins and the lack of intergenic regions between the genes suggest that these genes may also be members of this operon. While the association of *recC* with these genes is conserved throughout the examined members of the Pasteurellaceae, Q-PCR analysis does not support the inclusion of this gene within the regulon. While RecC may be involved in recombination of transforming DNA in *H. influenzae,* as has been demonstrated in *N. gonorrhoeae* (103), its vital role in other DNA metabolic activities would make it an unlikely candidate to be under tight control of competence development. Insertional mutagenesis of HI0939 demonstrates that its product, or that of the downstream genes, has a vital role in both uptake and binding of DNA. Unfortunately, complementation of the HI0939 mutant could not be accomplished since plasmid constructs bearing HI0939 could not be established in either the wild-type or mutant strains. While the translated products of these genes lack blastp matches outside of the Pasteurellaceae, BLAST searches against the conserved domain database (rpsblast) identifies matches of PulG for HI0938 and PulJ for HI0939. These proteins are type IV pseudopilins that form part of the type II secretion system for export of proteins from the periplasm in Gram-negative bacteria (117, 119). This system is believed to be composed of a pseudopilus that is related to and shares components with the Tfp biogenesis machinery. It has been proposed that this structure resembles a piston and functions to propel targeted proteins across the outer membrane in a manner similar to twitching

motility in the Tfp (71). In competence, this system could function either to export the DNA binding proteins to the cell surface or to facilitate the import of bound DNA through the outer membrane. The type II secreton has been best described for the secretion of pullulanase in *Klebsiella oxytoca* (117). The *H. influenzae* Rd KW20 genome contains homologs to several of the pullulanase operon genes, including *pulD* (HI0435 *comE*), *pulE* (HI0298 *pilB*), *pulF* (HI0297 *pilC*), *pulG* (HI0938), *pulJ* (HI0939) and *pulO* (HI0296 *pilD*). All of these genes appear to be transcribed from CRE-controlled operons and mutations in all of these operons result in abolition of DNA binding and uptake in *H. influenzae*, affirming the importance of the type II secreton to transformation (42, 61, 80, 149). I propose naming the products of HI0938 and HI0939 PulG and PulJ, respectively, due to their domain matches to these proteins.

The results of Q-PCR examination of *ligA* expression during competence development and the presence of putative CRE sites upstream of *ligA* homologs in *A. actinomycetemcomitans* and the *H. somnus* strains support the inclusion of *ligA* in the competence regulon. An earlier attempt by Preston *et al.* to create a mutant in HI1182 was unsuccessful leading to the conclusion that this gene might be essential (114). The success in deleting *ligA* in this study argues against the indispensability of this gene for survival. While clearly upregulated in response to competence development in Rd KW20, *ligA* does not appear to be indispensable to transformation. The lack of LigA homologs in many of the Pasteurellaceae is consistent with a non-critical role in transformation. It is possible that the NAD^+-dependent ligase (LigN) can substitute if LigA is absent. Interestingly, LigA appears to contain a signal peptide recognition sequence that would indicate a possible periplasmic location. LigA homologs are present

in several other bacterial species, including *Neisseria meningiditis* and *Vibrio cholerae*, and these homologs also contain putative signal peptides. It is difficult to conceive a function for a DNA ligase either as an exported or periplasmic protein either for DNA metabolism or for transformation.

Another finding of this work is that the three examined CRE sites located between divergently transcribed operons are responsible for control of transcription in a single direction. This may have important implications as eight of ten of the predicted CRE sites in the Rd KW20 sequence are located between divergently transcribed operons. In particular, CRE0439 is located between two operons in which insertions or deletions of genes within each operon have deleterious effects on transformation. Further examination of transcription from CRE0439 and the other CRE sites will answer the question of whether bidirectional control by CREs occurs or whether their frequent location between divergent operons is merely coincidental. Additionally, while CRE0364 is located closer to HI0364 than to HI0365, the element actually controls transcription of the HI0365 operon alone. Thus, the proximity of a promoter element to nearby genes is an inaccurate predictor of regulatory targeting.

In the second chapter of this work, the results of a project designed to allow the creation of a transposon mutagenesis library containing insertions in each of the Rd KW20 annotated genes were described. This library could then be used to create a bank of *H. influenzae* mutants that could be systematically screened to determine phenotypic changes related to each genetic disruption. While this approach was not successful in generating insertions in all of the Rd KW20 CDSs, the number of insertions recovered was sufficient to justify an attempt to fulfill the second goal of the project; *i.e.* to perform

systematic phenotypic screening with the mutant bank. The transposon insertions were transformed into Rd KW20 and antibiotic-resistant mutants were recovered and screened to determine their ability to be further transformed to novobiocin or streptomycin resistance using the static aerobic technique. The screening method utilized was able to identify a novel competence-related gene, HI1161. Sequence analysis confirmed that the transposon insertion was present in the HI1161 mutant strain (TMV1874-1). The HI1161 mutant resulted in a moderate decrease in transformation efficiency (approximately 30-fold decrease from wild-type cells). The transformation defect appears to be the result of decreased transcription of competence-regulated genes in HI1161 mutant strains when compared to Rd KW20. Additionally, *H. influenzae* mutant strains carrying Tn5 insertions in two genes immediately downstream of HI1161 are transformation proficient, thus the phenotype observed in TMV1874-1 is not due to any polar effects caused by the insertion.

Several possibilities could explain the regulatory defect observed in the HI1161 mutants. First, HI1161 could be directly involved in competence regulation by acting as a transcriptional regulator. However, computational analysis of the predicted product of HI1161 did not reveal any similarities to known transcriptional regulators. Second, HI1161 could be involved in modification of a signal important to competence induction. Finally, loss of HI1161 could result in pleiotropic effects that indirectly affect competence development. The defect in the HI1161 cannot be explained by decreased transcription of the putative competence regulatory gene *tfoX*. Additionally, it would appear that the cause is not related to a reduction in cAMP levels since the addition of exogenous cAMP was unable to complement the mutation. The product of HI1161 is

highly conserved among the eubacteria. In *B. subtilis,* the HI1161 homolog ComAB (51% identity, 64% similarity to HI1161) is cotranscribed with the early competence regulatory gene ComA; however, mutation of *comAB* did not affect competence development in that organism (157). HI1161 is included in the 4HBT superfamily (PF03061), members of which are predicted to be thioesterases and include various long-chain acyl-CoA thioester hydrolases. Interestingly, long-chain acyl-CoA molecules are involved in regulation of gene expression in prokaryotes (20); in particular, the expression of genes involved in fatty acid synthesis and degradation. Unfortunately, the lack of knowledge about HI1161 and its homologs in other bacteria precludes the determination, at this time, of the exact role that HI1161 might play in competence development.

In conclusion, the ability of comparative genomics to aid in the functional genomic analysis of *H. influenzae* has been demonstrated in this study. Characterization of the three newly identified competence-regulated operons by Q-PCR analysis indicated regulation consistent with that of other identified competence-regulated genes. Mutational analysis of HI0366 identified a novel transformation gene, *pilF2*, involved in uptake but not binding of DNA in *H. influenzae.* Mutational analysis of HI0939 identified a new gene, *pulJ,* possibly required for DNA binding and uptake in *H. influenzae*. The product of *pulJ* and HI0938 (*pulG*) are related to members of the type II secreton, confirming the importance of that system in transformation in *H. influenzae.* An ATP-dependent ligase encoded by HI1182/1183 (*ligA*) was shown to be upregulated in response to transfer into competence development media. However, mutational analysis indicated that LigA has only a minor role or that the loss of its function can be

complemented by the NAD-dependent ligase. Additionally, a large-scale global transposon insertion mutagenesis protocol led to the identification a novel gene, HI1161, for which disruption causes a moderate decrease in transformation ability. HI1161 has a role in moderating competence development that is not mediated by levels of cAMP.

REFERENCES (PART II)

1. **Aas, F. E., C. Lovold, and M. Koomey**. 2002. An inhibitor of DNA binding and uptake events dictates the proficiency of genetic transformation in *Neisseria gonorrhoeae*: mechanism of action and links to Type IV pilus expression. Mol. Microbiol. **46**:1441-1450.

2. **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**:749-760.

3. **Albano, M., R. Breitling, and D. A. Dubnau**. 1989. Nucleotide sequence and genetic organization of the *Bacillus subtilis comG* operon. J. Bacteriol. **171**:5386-5404.

4. **Alexander, H. and G. Leidy**. 1951. Determination of inherited traits of *Haemophilus influenzae* by desoxyribonucleic acid fractions isolated from type-specific cells. J. Exp. Med. **93**:345-359.

5. **Alonso, J. C., A. C. Stiege, and G. Luder**. 1993. Genetic recombination in *Bacillus subtilis* 168: effect of *recN, recF, recH* and *addAB* mutations on DNA repair and recombination. Mol. Gen. Genet. **239**:129-36.

6. **Alonso, J. C., R. H. Tailor, and G. Luder**. 1988. Characterization of recombination-deficient mutants of *Bacillus subtilis*. J. Bacteriol. **170**:3001-7.

7. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman**. 1990. Basic local alignment search tool. J. Mol. Biol. **215**:403-410.

8. **Anderson, D. G. and S. C. Kowalczykowski**. 1997. The recombination hot spot *chi* is a regulatory element that switches the polarity of DNA degradation by the RecBCD enzyme. Genes Dev. **11**:571-581.

9. **Anderson, D. G. and S. C. Kowalczykowski**. 1997. The translocating RecBCD enzyme stimulates recombination by directing RecA protein onto ssDNA in a *chi*-regulated manner. Cell **90**:77-86.

10. **Ando, T., D. A. Israel, K. Kusugami, and M. J. Blaser**. 1999. HP0333, a member of the *dprA* family, is involved in natural transformation in *Helicobacter pylori*. J. Bacteriol. **181**:5572-5580.

11. **Austin, S.** 1976. Wild-type and mutant *in vitro* products of an operon for ribonucleic acid polymerase subunits. J. Bacteriol. **127**:32-39.

12. **Averhoff, B. and A. Friedrich**. 2003. Type IV pili-related natural transformation systems: DNA transport in mesophilic and thermophilic bacteria. Arch. Microbiol. **180**:385-393.

13. **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**:4513-4518.

14. **Bannister, L. A.** 2000. An RNA secondary structure regulates *sxy* expression and competence development in *Haemophilus influenzae*. Ph.D. thesis. University of British Columbia.

15. **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.

16. **Barnhart, B. J. and S. H. Cox**. 1968. Radiation-sensitive and radiation-resistant mutants of *Haemophilus influenzae*. J. Bacteriol. **96**:280-282.

17. **Barouki, R. and H. O. Smith**. 1985. Reexamination of phenotypic defects in *rec-1* and *rec-2* mutants of *Haemophilus influenzae* Rd. J. Bacteriol. **163**:629-34.

18. **Barouki, R. and H. O. Smith**. 1986. Initial steps in *Haemophilus influenzae* transformation. Donor DNA binding in the *com10* mutant. J. Biol. Chem. **261**:8617-23.

19. **Berge, M., I. Mortier-Barriere, B. Martin, and J. P. Claverys**. 2003. Transformation of *Streptococcus pneumoniae* relies on DprA- and RecA-dependent protection of incoming DNA single strands. Mol. Microbiol. **50**:527-536.

20. **Black, P. N., N. J. Faergeman, and C. C. DiRusso**. 2000. Long-chain acyl-CoA-dependent regulation of gene expression in bacteria, yeast and mammals. J. Nutr. **130**:305S-309S.

21. **Botsford, J. L. and J. G. Harman**. 1992. Cyclic AMP in prokaryotes. Microbiol. Rev. **56**:100-122.

22. **Brooke, J. S. and M. A. Valvano**. 1996. Molecular cloning of the *Haemophilus influenzae gmhA* (*lpcA*) gene encoding a phosphoheptose isomerase required for lipooligosaccharide biosynthesis. J. Bacteriol. **178**:3339-3341.

23. **Bruckner, R. and F. Titgemeyer**. 2002. Carbon catabolite repression in bacteria: choice of the carbon source and autoregulatory limitation of sugar utilization. FEMS Microbiol. Lett. **209**:141-148.

24. **Cabedo, H., F. Macian, M. Villarroya, J. C. Escudero, M. Martinez-Vicente, E. Knecht, and M. E. Armengod**. 1999. The *Escherichia coli trmE* (*mnmE*)

gene, involved in tRNA modification, codes for an evolutionarily conserved GTPase with unusual biochemical properties. EMBO J. **18**:7063-7076.

25. **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.

26. **Chandler, M. S. and R. A. Smith**. 1996. Characterization of the *Haemophilus influenzae topA* locus: DNA topoisomerase I is required for genetic competence. Gene **169**:25-31.

27. **Chaussee, M. S. and S. A. Hill**. 1998. Formation of single-stranded DNA during DNA transformation of *Neisseria gonorrhoeae*. J. Bacteriol. **180**:5117-22.

28. **Chen, I. and D. Dubnau**. 2003. DNA transport during transformation. Front. Biosci. **8**:s544-s556.

29. **Chen, I. and D. Dubnau**. 2004. DNA uptake during bacterial transformation. Nat. Rev. Microbiol. **2**:241-249.

30. **Chen, I. and E. C. Gotschlich**. 2001. ComE, a competence protein from *Neisseria gonorrhoeae* with DNA-binding activity. J. Bacteriol. **183**:3160-3168.

31. **Chen, L. and M. F. Roberts**. 2000. Overexpression, purification, and analysis of complementation behavior of *E. coli* SuhB protein: comparison with bacterial and archaeal inositol monophosphatases. Biochemistry **39**:4145-4153.

32. **Cheng, C. and S. Shuman**. 1997. Characterization of an ATP-dependent DNA ligase encoded by *Haemophilus influenzae*. Nucleic Acids Res. **25**:1369-1374.

33. **Clifton, S. W., D. McCarthy, and B. A. Roe**. 1994. Sequence of the *rec-2* locus of *Haemophilus influenzae*: homologies to *comE*-ORF3 of *Bacillus subtilis* and *msbA* of *Escherichia coli*. Gene **146**:95-100.

34. **Concino, M. F. and S. H. Goodgal**. 1981. *Haemophilus influenzae* polypeptides involved in deoxyribonucleic acid uptake detected by cellular surface protein iodination. J. Bacteriol. **148**:220-31.

35. **Concino, M. F. and S. H. Goodgal**. 1982. DNA-binding vesicles released from the surface of a competence- deficient mutant of *Haemophilus influenzae*. J. Bacteriol. **152**:441-50.

36. **Dargis, M., P. Gourde, D. Beauchamp, B. Foiry, M. Jacques, and F. Malouin**. 1992. Modification in penicillin-binding proteins during *in vivo* development of genetic competence of *Haemophilus influenzae* is associated with a rapid change in the physiological state of cells. Infect. Immun. **60**:4024-4031.

37. **Deich, R. A. and L. C. Hoyer**. 1982. Generation and release of DNA-binding vesicles by *Haemophilus influenzae* during induction and loss of competence. J. Bacteriol. **152**:855-64.

38. **Deich, R. A. and H. O. Smith**. 1980. Mechanism of homospecific DNA uptake in *Haemophilus influenzae* transformation. Mol. Gen. Genet. **177**:369-74.

39. **Dixon, D. A. and S. C. Kowalczykowski**. 1993. The recombination hotspot chi is a regulatory sequence that acts by attenuating the nuclease activity of the *E. coli* RecBCD enzyme. Cell **73**:87-96.

40. **Donnenberg, M. S., H. Z. Zhang, and K. D. Stone**. 1997. Biogenesis of the bundle-forming pilus of enteropathogenic *Escherichia coli*: reconstitution of fimbriae in recombinant *E. coli* and role of DsbA in pilin stability--a review. Gene **192**:33-38.

41. **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.

42. **Dougherty, B. A. and H. O. Smith**. 1999. Identification of *Haemophilus influenzae* Rd transformation genes using cassette mutagenesis. Microbiology **145**:401-9.

43. **Dubnau, D.** 1999. DNA uptake in bacteria. Annu. Rev. Microbiol. **53**:217-44.

44. **Dubnau, D.** 2003. DNA uptake in *Bacillus subtilis*. American Society of Microbiology **103 General Meeting**.

45. **Facius, D. and T. F. Meyer**. 1993. A novel determinant (*comA*) essential for natural transformation competence in *Neisseria gonorrhoeae* and the effect of a *comA* defect on pilin variation. Mol. Microbiol. **10**:699-712.

46. **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.

47. **Forest, K. T. and J. A. Tainer**. 1997. Type-4 pilus-structure: outside to inside and top to bottom--a minireview. Gene **192**:165-169.

48. **Friedrich, A., C. Prust, T. Hartsch, A. Henne, and B. Averhoff**. 2002. Molecular analyses of the natural transformation machinery and identification of pilus structures in the extremely thermophilic bacterium *Thermus thermophilus* strain HB27. Appl. Environ. Microbiol. **68**:745-755.

49. **Gibbs, C. P., B. Y. Reimann, E. Schultz, A. Kaufmann, R. Haas, and T. F. Meyer**. 1989. Reassortment of pilin genes in *Neisseria gonorrhoeae* occurs by two distinct mechanisms. Nature **338**:651-652.

50. **Gish, W. and D. J. States**. 1993. Identification of protein coding regions by database similarity search. Nat. Genet. **3**:266-272.

51. **Goodgal, S. H.** 1982. DNA uptake in *Haemophilus* transformation. Annu. Rev. Genet. **16**:169-92.

52. **Goodgal, S. H. and R. M. Herriott**. 1961. Studies on transformations of *Hemophilus influenzae*. I. Competence. J. Gen. Physiol. **44**:1201-1227.

53. **Goodgal, S. H. and N. Notani**. 1968. Evidence that either strand of DNA can transform. J. Mol. Biol. **35**:449-53.

54. **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**:2184-2190.

55. **Graupner, S. and W. Wackernagel**. 2001. *Pseudomonas stutzeri* has two closely related *pilA* genes (Type IV pilus structural protein) with opposite influences on natural genetic transformation. J. Bacteriol. **183**:2359-2366.

56. **Graupner, S., N. Weger, M. Sohni, and W. Wackernagel**. 2001. Requirement of novel competence genes *pilT* and *pilU* of *Pseudomonas stutzeri* for natural transformation and suppression of *pilT* deficiency by a hexahistidine tag on the type IV pilus protein PilAI. J. Bacteriol. **183**:4694-4701.

57. **Gromkova, R. and S. Goodgal**. 1979. Transformation by plasmid and chromosomal DNAs in *Haemophilus parainfluenzae*. Biochem. Biophys. Res. Commun. **88**:1428-1434.

58. **Gromkova, R., P. Rowji, and H. Koornhof**. 1989. Induction of competence in Nonencapsulated and Encapsulated strains of *Haemophilus influenzae*. Curr. Microbiol. **19**:241-245.

59. **Gromkova, R. C., T. C. Mottalini, and M. G. Dove**. 1998. Genetic transformation in *Haemophilus parainfluenzae* clinical isolates. Curr. Microbiol. **37**:123-126.

60. **Gwinn, M. L., R. Ramanathan, H. O. Smith, and J. F. Tomb**. 1998. A new transformation-deficient mutant of *Haemophilus influenzae* Rd with normal DNA uptake. J. Bacteriol. **180**:746-8.

61. **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.

62. **Gwinn, M. L., D. Yi, H. O. Smith, and J. F. Tomb**. 1996. Role of the two-component signal transduction and the phosphoenolpyruvate: carbohydrate phosphotransferase systems in competence development of *Haemophilus influenzae* Rd. J. Bacteriol. **178**:6366-8.

63. **Hagervall, T. G., S. C. Pomerantz, and J. A. McCloskey**. 1998. Reduced misreading of asparagine codons by *Escherichia coli* tRNALys with hypomodified derivatives of 5-methylaminomethyl-2-thiouridine in the wobble position. J. Mol. Biol. **284**:33-42.

64. **Haijema, B. J., G. Venema, and J. Kooistra**. 1996. The C terminus of the AddA subunit of the *Bacillus subtilis* ATP-dependent DNase is required for the ATP-dependent exonuclease activity but not for the helicase activity. J. Bacteriol. **178**:5086-5091.

65. **Hamoen, L. W., G. Venema, and O. P. Kuipers**. 2003. Controlling competence in *Bacillus subtilis*: shared use of regulators. Microbiology **149**:9-17.

66. **Heckels, J. E.** 1989. Structure and function of pili of pathogenic *Neisseria* species. Clin. Microbiol. Rev. **2 Suppl**:S66-S73.

67. **Hegde, S. P., M. H. Qin, X. H. Li, M. A. Atkinson, A. J. Clark, M. Rajagopalan, and M. V. Madiraju**. 1996. Interactions of RecF protein with RecO, RecR, and single-stranded DNA binding proteins reveal roles for the RecF-RecO-RecR complex in DNA repair and recombination. Proc. Natl. Acad. Sci. U. S. A. **93**:14468-73.

68. **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.

69. **Higgins, C. F., C. J. Dorman, D. A. Stirling, L. Waddell, I. R. Booth, G. May, and E. Bremer**. 1988. A physiological role for DNA supercoiling in the osmotic regulation of gene expression in *S. typhimurium* and *E. coli*. Cell **52**:569-84.

70. **Hiltke, T. J., A. T. Schiffmacher, A. J. Dagonese, S. Sethi, and T. F. Murphy**. 2003. Horizontal transfer of the gene encoding outer membrane protein P2 of nontypeable *Haemophilus influenzae*, in a patient with chronic obstructive pulmonary disease. J. Infect. Dis. **188**:114-117.

71. **Hobbs, M. and J. S. Mattick**. 1993. Common components in the assembly of type 4 fimbriae, DNA transfer systems, filamentous phage and protein-secretion apparatus: a general system for the formation of surface-associated protein complexes. Mol. Microbiol. **10**:233-243.

72. **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**:3045-3051.

73. **Jensen, P. R. and O. Michelsen**. 1992. Carbon and energy metabolism of *atp* mutants of *Escherichia coli*. J. Bacteriol. **174**:7635-7641.

74. **Jett, B. D., K. L. Hatter, M. M. Huycke, and M. S. Gilmore**. 1997. Simplified agar plate method for quantifying viable bacteria. Biotechniques **23**:648-650.

75. **Kahn, M., M. Concino, R. Gromkova, and S. Goodgal**. 1979. DNA binding activity of vesicles produced by competence deficient mutants of *Haemophilus*. Biochem. Biophys. Res. Commun. **87**:764-72.

76. **Kahn, M. E., F. Barany, and H. O. Smith**. 1983. Transformasomes: specialized membranous structures that protect DNA during *Haemophilus* transformation. Proc. Natl. Acad. Sci. U. S. A. **80**:6927-31.

77. **Kahn, M. E., G. Maul, and S. H. Goodgal**. 1982. Possible mechanism for donor DNA binding and transport in *Haemophilus*. Proc. Natl. Acad. Sci. U. S. A. **79**:6370-4.

78. **Kahn, M. E. and H. O. Smith**. 1984. Transformation in *Haemophilus*: a problem in membrane biology. J. Membr. Biol. **81**:89-103.

79. **Kang, Y., H. Liu, S. Genin, M. A. Schell, and T. P. Denny**. 2002. *Ralstonia solanacearum* requires type 4 pili to adhere to multiple surfaces and for natural transformation and virulence. Mol. Microbiol. **46**:427-437.

80. **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.

81. **Karudapuram, S., X. Zhao, and G. J. Barcak**. 1995. DNA sequence and characterization of *Haemophilus influenzae dprA+*, a gene required for chromosomal but not plasmid DNA transformation. J. Bacteriol. **177**:3235-40.

82. **Kowalczykowski, S. C., D. A. Dixon, A. K. Eggleston, S. D. Lauder, and W. M. Rehrauer**. 1994. Biochemistry of homologous recombination in *Escherichia coli*. Microbiol. Rev. **58**:401-65.

83. **Kroll, J. S., B. M. Loynds, and P. R. Langford**. 1992. Palindromic *Haemophilus* DNA uptake sequences in presumed transcriptional terminators from *H. influenzae* and *H. parainfluenzae*. Gene **114**:151-2.

84. **Kupfer, D. M. and D. McCarthy**. 1992. *rec-2*-dependent phage recombination in *Haemophilus influenzae*. J. Bacteriol. **174**:4960-6.

85. **Larribe, M., M. K. Taha, A. Topilko, and C. Marchal**. 1997. Control of *Neisseria gonorrhoeae* pilin gene expression by environmental factors: involvement of the *pilA/pilB* regulatory genes. Microbiology **143 (Pt 5)**:1757-1764.

86. **LeClerc, J. E. and J. K. Setlow**. 1975. Single-strand regions in the deoxyribonucleic acid of competent *Haemophilus influenzae*. J. Bacteriol. **122**:1091-102.

87. **Lee, M. S., B. A. Dougherty, A. C. Madeo, and D. A. Morrison**. 1999. Construction and analysis of a library for random insertional mutagenesis in *Streptococcus pneumoniae*: use for recovery of mutants defective in genetic transformation and for identification of essential genes. Appl. Environ. Microbiol. **65**:1883-1890.

88. **Lim, K., A. Tempczyk, J. F. Parsons, N. Bonander, J. Toedt, Z. Kelman, A. Howard, E. Eisenstein, and O. Herzberg**. 2003. Crystal structure of YbaB from *Haemophilus influenzae* (HI0442), a protein of unknown function coexpressed with the recombinational DNA repair protein RecR. Proteins **50**:375-379.

89. **Livak, K. J. and T. D. Schmittgen**. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(ΔΔCT) method. Methods **25**:402-408.

90. **Londono-Vallejo, J. A. and D. Dubnau**. 1994. Mutation of the putative nucleotide binding site of the *Bacillus subtilis* membrane protein ComFA abolishes the uptake of DNA during transformation. J. Bacteriol. **176**:4642-4645.

91. **Long, C. D., S. F. Hayes, J. P. van Putten, H. A. Harvey, M. A. Apicella, and H. S. Seifert**. 2001. Modulation of gonococcal piliation by regulatable transcription of *pilE*. J. Bacteriol. **183**:1600-1609.

92. **Ma, C. and R. J. Redfield**. 2000. Point mutations in a peptidoglycan biosynthesis gene cause competence induction in *Haemophilus influenzae*. J. Bacteriol. **182**:3323-30.

93. **MacFadyen, L. P.** 1999. Regulation of intracellular cAMP levels and competence development in *Haemophilus influenzae* by a phosphoenolpyruvate: fructose phospotransferase system. Ph.D. thesis. University of British Columbia.

94. **MacFadyen, L. P.** 2000. Regulation of competence development in *Haemophilus influenzae*. J. Theor. Biol. **207**:349-59.

95. **MacFadyen, L. P., D. Chen, H. C. Vo, D. Liao, R. Sinotte, and R. J. Redfield**. 2001. Competence development by *Haemophilus influenzae* is regulated by the availability of nucleic acid precursors. Mol. Microbiol. **40**:700-7.

96. **MacFadyen, L. P., I. R. Dorocicz, J. Reizer, M. H. Saier, Jr., and R. J. Redfield**. 1996. Regulation of competence development and sugar utilization in *Haemophilus influenzae* Rd by a phosphoenolpyruvate:fructose phosphotransferase system. Mol. Microbiol. **21**:941-52.

97. **MacFadyen, L. P., C. Ma, and R. J. Redfield**. 1998. A 3',5' cyclic AMP (cAMP) phosphodiesterase modulates cAMP levels and optimizes competence in *Haemophilus influenzae* Rd. J. Bacteriol. **180**:4401-5.

98. **Maruyama, I. N., A. H. Yamamoto, and Y. Hirota**. 1988. Determination of gene products and coding regions from the *murE-murF* region of *Escherichia coli*. J. Bacteriol. **170**:3786-3788.

99. **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**:1159-61.

100. **McCarthy, D.** 1982. Plasmid recombination in *Haemophilus influenzae*. J. Mol. Biol. **157**:577-96.

101. **McCarthy, D.** 1989. Cloning of the *rec-2* locus of *Haemophilus influenzae*. Gene **75**:135-43.

102. **McCarthy, D. and D. M. Kupfer**. 1987. Electron microscopy of single-stranded structures in the DNA of competent *Haemophilus influenzae* cells. J. Bacteriol. **169**:565-71.

103. **Mehr, I. J. and H. S. Seifert**. 1998. Differential roles of homologous recombination pathways in *Neisseria gonorrhoeae* pilin antigenic variation, DNA transformation and DNA repair. Mol. Microbiol. **30**:697-710.

104. **Meima, R., C. Eschevins, S. Fillinger, A. Bolhuis, L. W. Hamoen, R. Dorenbos, W. J. Quax, J. M. van Dijl, R. Provvedi, I. Chen, D. Dubnau, and S. Bron**. 2002. The *bdbDC* operon of *Bacillus subtilis* encodes thiol-disulfide oxidoreductases required for competence development. J. Biol. Chem. **277**:6994-7001.

105. **Mitchell, M. A., K. Skowronek, L. Kauc, and S. H. Goodgal**. 1991. Electroporation of *Haemophilus influenzae* is effective for transformation of plasmid but not chromosomal DNA. Nucleic Acids Res. **19**:3625-8.

106. **Nielsen, H., J. Engelbrecht, S. Brunak, and G. Von Heijne**. 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng. **10**:1-6.

107. **Notani, N. K. and J. K. Setlow**. 1974. Mechanism of bacterial transformation and transfection. Prog. Nucleic Acid Res. Mol. Biol. **14**:39-100.

108. **Ogura, M., H. Yamaguchi, K. Kobayashi, N. Ogasawara, Y. Fujita, and T. Tanaka**. 2002. Whole-genome analysis of genes regulated by the *Bacillus subtilis* competence transcription factor ComK. J. Bacteriol. **184**:2344-2351.

109. **Pelaez, A. I., R. M. Ribas-Aparicio, A. Gomez, and M. R. Rodicio**. 2001. Structural and functional characterization of the *recR* gene of *Streptomyces*. Mol. Genet. Genomics **265**:663-672.

110. **Pestova, E. V., L. S. Havarstein, and D. A. Morrison**. 1996. Regulation of competence for genetic transformation in *Streptococcus pneumoniae* by an auto-induced peptide pheromone and a two-component regulatory system. Mol. Microbiol. **21**:853-862.

111. **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**:2701-2710.

112. **Poje, G. and R. J. Redfield**. 2003. Transformation of *Haemophilus influenzae*. Methods Mol. Med. **71**:57-70.

113. **Porstendorfer, D., U. Drotschmann, and B. Averhoff**. 1997. A novel competence gene, *comP*, is essential for natural transformation of *Acinetobacter* sp. strain BD413. Appl. Environ. Microbiol. **63**:4150-4157.

114. **Preston, A., D. Maskell, A. Johnson, and E. R. Moxon**. 1996. Altered lipopolysaccharide characteristic of the *I69* phenotype in *Haemophilus influenzae* results from mutations in a novel gene, *isn*. J. Bacteriol. **178**:396-402.

115. **Provvedi, R., I. Chen, and D. Dubnau**. 2001. NucA is required for DNA cleavage during transformation of *Bacillus subtilis*. Mol. Microbiol. **40**:634-644.

116. **Provvedi, R. and D. Dubnau**. 1999. ComEA is a DNA receptor for transformation of competent *Bacillus subtilis*. Mol. Microbiol. **31**:271-280.

117. **Pugsley, A. P.** 1993. The complete general secretory pathway in gram-negative bacteria. Microbiol. Rev. **57**:50-108.

118. **Pugsley, A. P., N. Bayan, and N. Sauvonnet**. 2001. Disulfide bond formation in secreton component PulK provides a possible explanation for the role of DsbA in pullulanase secretion. J. Bacteriol. **183**:1312-1319.

119. **Pugsley, A. P., O. Francetic, O. M. Possot, N. Sauvonnet, and K. R. Hardie**. 1997. Recent progress and future directions in studies of the main terminal branch of the general secretory pathway in Gram-negative bacteria--a review. Gene **192**:13-19.

120. **Ranhand, J. M. and H. C. Lichstein**. 1969. Effect of selected antibiotics and other inhibitors on competence development in *Haemophilus influenzae*. J. Gen. Microbiol. **55**:37-43.

121. **Redfield, R. J.** 1991. *sxy-1*, a *Haemophilus influenzae* mutation causing greatly enhanced spontaneous competence. J. Bacteriol. **173**:5612-8.

122. **Redfield, R. J.** 2001. Competence genes summary (Includes unpublished data). http://www.zoology.ubc.ca/~redfield/CIHR/HIgenes.pdf .

123. **Roca, A. I. and M. M. Cox**. 1997. RecA protein: structure, function, and role in recombinational DNA repair. Prog. Nucleic Acid Res. Mol. Biol. **56**:129-223.

124. **Rolfes, R. J. and H. Zalkin**. 1988. Regulation of *Escherichia coli purF*. Mutations that define the promoter, operator, and purine repressor gene. J. Biol. Chem. **263**:19649-19652.

125. **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**:7986-7990.

126. **Rutherford, K., J. Parkhill, J. Crook, T. Horsnell, P. Rice, M. A. Rajandream, and B. Barrell**. 2000. Artemis: sequence visualization and annotation. Bioinformatics **16**:944-945.

127. **Sanchez, L., W. Pan, M. Vinas, and H. Nikaido**. 1997. The *acrAB* homolog of *Haemophilus influenzae* codes for a functional multidrug efflux pump. J. Bacteriol. **179**:6855-6857.

128. **Schlor, S., M. Herbert, M. Rodenburg, J. Blass, and J. Reidl**. 2000. Characterization of ferrochelatase (*hemH*) mutations in *Haemophilus influenzae*. Infect. Immun. **68**:3007-3009.

129. **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.

130. **Setlow, J. K., M. E. Boling, D. P. Allison, and K. L. Beattie**. 1973. Relationship between prophage induction and transformation in *Haemophilus influenzae*. J. Bacteriol. **115**:153-61.

131. **Setlow, J. K., M. E. Boling, K. L. Beattie, and R. F. Kimball**. 1972. A complex of recombination and repair genes in *Haemophilus influenzae*. J. Mol. Biol. **68**:361-378.

132. **Setlow, J. K., D. Spikes, and K. Griffin**. 1988. Characterization of the *rec-1* gene of *Haemophilus influenzae* and behavior of the gene in *Escherichia coli*. J. Bacteriol. **170**:3876-81.

133. **Shan, Q., J. M. Bork, B. L. Webb, R. B. Inman, and M. M. Cox**. 1997. RecA protein filaments: end-dependent dissociation from ssDNA and stabilization by RecO and RecR proteins. J. Mol. Biol. **265**:519-40.

134. **Sharma, R. C. and R. T. Schimke**. 1996. Preparation of electrocompetent *E. coli* using salt-free growth medium. Biotechniques **20**:42-44.

135. **Smeets, L. C., J. J. Bijlsma, E. J. Kuipers, C. M. Vandenbroucke-Grauls, and J. G. Kusters**. 2000. The *dprA* gene is required for natural transformation of *Helicobacter pylori*. FEMS Immunol. Med Microbiol. **27**:99-102.

136. **Smeets, L. C. and J. G. Kusters**. 2002. Natural transformation in *Helicobacter pylori*: DNA transport in an unexpected way. Trends Microbiol. **10**:159-162.

137. **Smith, H. O., D. B. Danner, and R. A. Deich**. 1981. Genetic transformation. Annu. Rev. Biochem. **50**:41-68.

138. **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.

139. **Solomon, J. M., R. Magnuson, A. Srivastava, and A. D. Grossman**. 1995. Convergent sensing pathways mediate response to two extracellular competence factors in *Bacillus subtilis*. Genes Dev. **9**:547-558.

140. **Sparling, P. F.** 1966. Genetic transformation of *Neisseria gonorrhoeae* to streptomycin resistance. J. Bacteriol. **92**:1364-1371.

141. **Spencer, H. T. and R. M. Herriott**. 1965. Development of competence of *Haemophilus influenzae*. J. Bacteriol. **90**:911-920.

142. **Steck, T. R., R. J. Franco, J. Y. Wang, and K. Drlica**. 1993. Topoisomerase mutations affect the relative abundance of many *Escherichia coli* proteins. Mol. Microbiol. **10**:473-81.

143. **Stone, B. J. and Y. A. Kwaik**. 1999. Natural competence for DNA transformation by *Legionella pneumophila* and its association with expression of type IV pili. J. Bacteriol. **181**:1395-1402.

144. **Stuy, J. H.** 1989. Cloning and characterization of the *Haemophilus influenzae* Rd *rec-1*+ gene. J. Bacteriol. **171**:4395-401.

145. **Sutrina, S. L. and J. J. Scocca**. 1976. Phospholipids of *Haemophilus influenzae* Rd during exponential growth and following the development of competence for genetic transformation. J. Gen. Microbiol. **92**:410-2.

146. **Tartof, K. D. and C. A. Hobbs**. 1988. New cloning vectors and techniques for easy and rapid restriction mapping. Gene **67**:169-182.

147. **Tomb, J. F.** 1992. A periplasmic protein disulfide oxidoreductase is required for transformation of *Haemophilus influenzae* Rd. Proc. Natl. Acad. Sci. U. S. A. **89**:10252-10256.

148. **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.

149. **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.

150. **Tortosa, P. and D. Dubnau**. 1999. Competence for transformation: a matter of taste. Curr. Opin. Microbiol. **2**:588-592.

151. **Urbonavicius, J., Q. Qian, J. M. Durand, T. G. Hagervall, and G. R. Bjork**. 2001. Improvement of reading frame maintenance is a common function for several tRNA modifications. EMBO J. **20**:4863-4873.

152. **Wachi, M., M. Doi, Y. Okada, and M. Matsuhashi**. 1989. New mre genes *mreC* and *mreD*, responsible for formation of the rod shape of *Escherichia coli* cells. J. Bacteriol. **171**:6511-6.

153. **Wachi, M., M. Doi, S. Tamaki, W. Park, S. Nakajima-Iijima, and M. Matsuhashi**. 1987. Mutant isolation and molecular cloning of mre genes, which determine cell shape, sensitivity to mecillinam, and amount of penicillin-binding proteins in *Escherichia coli*. J. Bacteriol. **169**:4935-40.

154. **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**:3442-3449.

155. **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**:249-255.

156. **Watson, A. A., R. A. Alm, and J. S. Mattick**. 1996. Identification of a gene, *pilF*, required for type 4 fimbrial biogenesis and twitching motility in *Pseudomonas aeruginosa*. Gene **180**:49-56.

157. **Weinrauch, Y., N. Guillen, and D. A. Dubnau**. 1989. Sequence and transcription mapping of *Bacillus subtilis* competence genes *comB* and *comA*, one of which is related to a family of bacterial regulatory determinants. J. Bacteriol. **171**:5362-5375.

158. **Whitby, P. W., D. J. Morton, and T. L. Stull**. 1998. Construction of antibiotic resistance cassettes with multiple paired restriction sites for insertional mutagenesis of *Haemophilus influenzae*. FEMS Microbiol. Lett. **158**:57-60.

159. **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.

160. **Wise, E. M., Jr., S. P. Alexander, and M. Powers**. 1973. Adenosine 3':5'-cyclic monophosphate as a regulator of bacterial transformation. Proc. Natl. Acad. Sci. U. S. A. **70**:471-474.

161. **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**:321-330.

162. **Wolfgang, M., J. P. van Putten, S. F. Hayes, D. Dorward, and M. Koomey**. 2000. Components and dynamics of fiber formation define a ubiquitous biogenesis pathway for bacterial pili. EMBO J. **19**:6408-6418.

163. **Wolfgang, M., J. P. van Putten, S. F. Hayes, and M. Koomey**. 1999. The *comP* locus of *Neisseria gonorrhoeae* encodes a type IV prepilin that is dispensable for pilus biogenesis but essential for natural transformation. Mol. Microbiol. **31**:1345-1357.

164. **Yoshihara, S., X. Geng, S. Okamoto, K. Yura, T. Murata, M. Go, M. Ohmori, and M. Ikeuchi**. 2001. Mutational analysis of genes involved in pilus structure, motility and transformation competency in the unicellular motile cyanobacterium *Synechocystis* sp. PCC 6803. Plant Cell Physiol. **42**:63-73.

165. **Zalkin, H. and P. Nygaard**. 1996. Biosynthesis of purine nucleotides, p. 561-579. *In* F. C. Neidhardt (ed.), *Escherichia coli* and *Salmonella* cellular and molecular biology. ASM Press, Washington, D.C.

166. **Zoon, K. C., M. Habersat, and J. J. Scocca**. 1976. Synthesis of envelope polypeptides by *Haemophilus influenzae* during development of competence for genetic transformation. J. Bacteriol. **127**:545-54.

167. **Zoon, K. C. and J. J. Scocca**. 1975. Constitution of the cell envelope of *Haemophilus influenzae* in relation to competence for genetic transformation. J. Bacteriol. **123**:666-77.

168. **Zulty, J. J. and G. J. Barcak**. 1993. Structural organization, nucleotide sequence, and regulation of the *Haemophilus influenzae rec-1*+ gene. J. Bacteriol. **175**:7269-81.

169. **Zulty, J. J. and G. J. Barcak**. 1995. Identification of a DNA transformation gene required for *com101A*+ expression and supertransformer phenotype in *Haemophilus influenzae*. Proc. Natl. Acad. Sci. U. S. A. **92**:3616-20.

APPENDICES

Appendix A – Mapped Tn7 insertions (pASC13 libraries)

Allele/Plasmid	CDS/Vector	Protein/Gene	Mutation coordinate	Disrupted codon	Total codons	Library	Comments
pASC0252	pGPS1	donor plasmid	2794			RdShear	
pASC0252	HI0509	1,4-dihydroxy-2-naphthoate octaprenyltransferase (menA)	526082	18	308	RdShear	
pASC0252	HI0509	1,4-dihydroxy-2-naphthoate octaprenyltransferase (menA)	526085	17	308	RdShear	
pASC0253	HI0036	ABC transporter, ATP-binding protein	39023	262	592	RdShear	
pASC0254	HI0491	conserved hypothetical protein	509785	111	167	RdShear	
pASC0255	HI0904	prolipoprotein diacylglyceryl transferase (lgt)	958731	12	268	RdShear	
pASC0256	HI0904	prolipoprotein diacylglyceryl transferase (lgt)	958731	12	268	RdShear	
pASC0261	pGPS1	donor plasmid	1100			RdShear	
pASC0261	HI0872	undecaprenyl-phosphate galactosephosphotransferase (rfbP)	924322	38	471	RdShear	
pASC0263	HI1480	H. influenzae predicted coding region HI1480	1563952	18	156	RdShear	
pASC0264	pGPS1	donor plasmid	1100			RdShear	
pASC0264	HI0520	conserved hypothetical protein	543299	133	262	RdShear	
pASC0265	HI1302	asparaginyl-tRNA-synthetase (asnS)	1382561	151	477	RdShear	
pASC0266	HI0184	esterase	198500	138	275	RdShear	Intergenic
pASC0268	HI0544	ribosomal protein L9 (rpL9)	567778	146	149	RdShear	
pASC0270	pGPS1	donor plasmid	1100			RdShear	
pASC0270	HI0364	penicillin-binding protein 7, putative	387350	298	292	RdShear	Intergenic
pASC0272	HI0979	nitrogen fixation protein (nifR3)	1038025	301	330	RdShear	
pASC0273	pGPS1	donor plasmid	2794			RdShear	
pASC0273	HI0364	penicillin-binding protein 7, putative	387347	301	292	RdShear	Intergenic
pASC0275	pGPS1	donor plasmid	1100			RdShear	
pASC0275	HI0364	penicillin-binding protein 7, putative	387350	298	292	RdShear	Intergenic
pASC0276	HI0175	conserved hypothetical protein	188516	119	244	RdShear	
pASC0276	HI0184	esterase	198499	139	275	RdShear	Intergenic
pASC0276	HI0184	esterase	198500	138	275	RdShear	Intergenic
pASC0278	HI0247	mutated adhesion/penetration protease	274796	99	1408	RdShear	Intergenic

pASC0279	HI0812	glucosephosphate uridylyltransferase (galU)	862249	93	295	RdShear	Intergenic
pASC0280	HI1193	branched-chain-amino-acid transaminase (ilvE)	1259594	13	343	RdShear	
pASC0282	HI0925	conserved hypothetical protein	984119	82	121	RdShear	
pASC0284	HI0812	glucosephosphate uridylyltransferase (galU)	862249	93	295	RdShear	
pASC0285	HI0162	lipoprotein, putative	178864	13	199	RdShear	
pASC0287	HI1462	conserved hypothetical protein	1544562	160	454	RdShear	
pASC0288	HI0189	glutamate dehydrogenase (gdhA)	203714	141	449	RdShear	
pASC0289	HI1189	conserved hypothetical protein	1256090	426	211	RdShear	Intergenic
pASC0290	pGPS1	donor plasmid	1100			RdShear	
pASC0290	HI0573	slyX protein (slyX)	593699	19	73	RdShear	
pASC0291	HI1258	transcription-repair coupling factor (mfd)	1332325	48	1146	RdShear	
pASC0291	HI1636	phosphoenolpyruvate carboxylase (ppc)	1700759	651	879	RdShear	
pASC0292	HI1389	anthanilate phosphoribosyltransferase (trpD)	1485588	140	333	RdShear	
pASC0293	HI0232	conserved hypothetical protein	263727	45	462	RdShear	
pASC0295	HI0392	O-antigen acetylase, putative	414004	227	245	RdShear	Intergenic
pASC0299	pASC13	vector	1298			RdShear	
pASC02x0	HI0090	conserved hypothetical protein	97543	115	237	RdShear	
pASC0300	pASC13	vector	1298			RdShear	
pASC0301	pGPS1	donor plasmid	1100			RdShear	
pASC0301	pASC13	vector	1305			RdShear	
pASC0302	pASC13	vector	1220			RdShear	
pASC0303	pASC13	vector	1298			RdShear	
pASC0304	pASC13	vector	1298			RdShear	
pASC0305	pASC13	vector	1298			RdShear	
pASC0306	pASC13	vector	1298			RdShear	
pASC0307	pASC13	vector	1298			RdShear	
pASC0336	HI1466.1	H. influenzae predicted coding region HI1466.1	1552440	337	345	RdShear	
pASC0336	HI1467	ABC transporter, ATP-binding protein	1552583	19	589	RdShear	
pASC0336	HI1470	iron chelatin ABC transporter, ATP-binding protein, putative	1555693	36	253	RdShear	
pASC0337	pGPS1	donor plasmid	1100			RdShear	
pASC0337	HI0344	napA protein frameshift	371509	365	832	RdShear	
pASC0338	HI1467	ABC transporter, ATP-binding protein	1553355	276	589	RdShear	
pASC0340	HI0366	fimbrial biogenesis and twitching motility protein, putative	390175	49	179	RdShear	

pASC0340	HI0367	conserved hypothetical protein	391048	137	303	RdShear	
pASC0341	pASC13	vector	515			RdShear	
pASC0342	pGPS1	donor plasmid	2796			RdShear	
pASC0342	HI1468	ribosomal protein S15 (rpS15)	1554557	48	89	RdShear	
pASC0343	HI0936	cytochrome C-type biogenesis	995305	280	635	RdShear	
pASC0344	pGPS1	donor plasmid	1100			RdShear	
pASC0344	HI0344	napA protein frameshift	371628	405	832	RdShear	
pASC0347	pASC13	vector	2287			RdShear	
pASC0380	HI0365	conserved hypothetical protein	389460	236	390	RdShear	
pASC0383	HI0344	napA protein frameshift	372247	611	832	RdShear	
pASC0388	pGPS1	donor plasmid	1100			RdShear	
pASC0388	HI0344	napA protein frameshift	372178	588	832	RdShear	

Gray shading indicates multiple insertions mapped in the same sequencing reaction

Appendix B – Mapped Tn7 insertions (pASC15 libraries)

Allele/Plasmid	CDS/Vector	Protein/Gene	Mutation coordinate	Disrupted codon	Total codons	Library	Comments
pASC0900	HI0936	cytochrome C-type biogenesis	995057	362	635	*Pvu*II	
pASC0905	HI0936	cytochrome C-type biogenesis	995057	362	635	*Pvu*II	
pASC1100-12	HI0936	cytochrome C-type biogenesis	995057	362	635	RdShear	
pASC1136-55	HI0936	cytochrome C-type biogenesis	995057	362	635	RdShear	
pASC1186	HI0936	cytochrome C-type biogenesis	995622	174	635	*Fsp*I	
pASC1187	HI0344	napA protein frameshift	371569	385	832	*Fsp*I	
pASC1188	HI0936	cytochrome C-type biogenesis	995519	208	635	*Fsp*I	
pASC1189	HI0344	napA protein frameshift	371759	448	832	*Fsp*I	
pASC1190	HI1470	iron chelatin ABC trnsp, ATP-binding protein	1555347	152	253	*Fsp*I	
pASC1191	HI0344	napA protein frameshift	372253	613	832	*Fsp*I	
pASC1192	HI0344	napA protein frameshift	371936	507	832	*Fsp*I	
pASC1193	pASC15	vector	1788			*Fsp*I	
pASC1193	HI0344	napA protein frameshift	371715	434	832	*Fsp*I	
pASC1194	pASC15	vector	1788			*Fsp*I	
pASC1194	HI0935	thiol:disulfide interchange protein (dsbE)	993916	107	176	*Fsp*I	
pASC1195	HI0344	napA protein frameshift	371864	483	832	*Fsp*I	
pASC1196	HI0344	napA protein frameshift	371831	472	832	*Fsp*I	
pASC1197	HI0344	napA protein frameshift	371557	381	832	*Fsp*I	
pASC1198	HI0344	napA protein frameshift	372378	655	832	*Fsp*I	
pASC1199	pASC15	vector	128			*Fsp*I	
pASC1200	HI0344	napA protein frameshift	372402	663	832	*Fsp*I	
pASC1201	pASC15	vector	1788			*Fsp*I	
pASC1201	HI0344	napA protein frameshift	372215	600	832	*Fsp*I	
pASC1202	HI0936	cytochrome C-type biogenesis	995883	87	635	*Fsp*I	

pASC1203	HI0344	napA protein frameshift	371573	386	832 *Fsp*I	
pASC1204	HI0935	thiol:disulfide interchange protein (dsbE)	993969	89	176 *Fsp*I	
pASC1205	HI0936	cytochrome C-type biogenesis	995040	368	635 *Fsp*I	
pASC1206	pASC15	vector	456		*Fsp*I	*Hinf* insert @ 372534
pASC1207	HI1467	ABC transporter, ATP-binding protein	1553522	332	589 *Fsp*I	
pASC1208	HI0935	thiol:disulfide interchange protein (dsbE)	994065	57	176 *Fsp*I	
pASC1209	pASC15		1788		*Fsp*I	
pASC1209	HI1468	ribosomal protein S15 (rpS15)	1554308	124	89 *Fsp*I	Intergenic
pASC1210	HI1470	iron chelatin ABC trnsp, ATP-binding protein	1555166	212	253 *Fsp*I	
pASC1211	pASC15	vector	240		*Fsp*I	
pASC1212	pASC15	vector	1788		*Fsp*I	
pASC1212	HI0344	napA protein frameshift	372130	572	832 *Fsp*I	
pASC1213	pASC15	vector	2221		*Fsp*I	
pASC1214	pASC15		1788		*Fsp*I	
pASC1214	HI1469	H. influenzae predicted coding region HI1469	1554884	49	115 *Fsp*I	
pASC1215	HI0344	napA protein frameshift	371561	382	832 *Fsp*I	
pASC1400	HI0936	cytochrome C-type biogenesis	995057	362	635 *Pvu*II	
pASC1405	HI0936	cytochrome C-type biogenesis	995057	362	635 *Pvu*II	
pASC1410	HI0936	cytochrome C-type biogenesis	995057	362	635 *Pvu*II	
pASC1415	HI0936	cytochrome C-type biogenesis	995057	362	635 *Pvu*II	
pASC1440	HI0936	cytochrome C-type biogenesis	995057	362	635 *Pvu*II	
pASC1445	HI0936	cytochrome C-type biogenesis	995057	362	635 *Pvu*II	
pASC1449	HI0936	cytochrome C-type biogenesis	995057	362	635 *Pvu*II	
pASC1455	HI0936	cytochrome C-type biogenesis	995057	362	635 *Pvu*II	
pASC1460	HI0936	cytochrome C-type biogenesis	995057	362	635 *Pvu*II	
pASC1465	HI0936	cytochrome C-type biogenesis	995057	362	635 *Pvu*II	
pASC1470	HI0936	cytochrome C-type biogenesis	995057	362	635 *Pvu*II	
pASC1471	HI0594	conserved hypothetical transmembrane protein	615416	356	509 *Pvu*II	Intergenic
pASC1472	pASC15	vector	119		*Pvu*II	
pASC1473	HI0936	cytochrome C-type biogenesis	995057	362	635 *Pvu*II	
pASC1474	HI0936	cytochrome C-type biogenesis	995057	362	635 *Pvu*II	
pASC1475	HI0936	cytochrome C-type biogenesis	995057	362	635 *Pvu*II	

pASC1476	HI0936	cytochrome C-type biogenesis	995057	362	635 *Pvu*II
pASC1477	HI0936	cytochrome C-type biogenesis	995057	362	635 *Pvu*II
pASC1478	HI0936	cytochrome C-type biogenesis	995057	362	635 *Pvu*II
pASC1479	HI0936	cytochrome C-type biogenesis	995057	362	635 *Pvu*II
pASC1480	pASC15	vector	119		*Pvu*II
pASC1481	HI0936	cytochrome C-type biogenesis	995057	362	635 *Pvu*II
pASC1482	HI0936	cytochrome C-type biogenesis	995057	362	635 *Pvu*II
pASC1483	HI0936	cytochrome C-type biogenesis	995057	362	635 *Pvu*II
pASC1484	HI0936	cytochrome C-type biogenesis	995057	362	635 *Pvu*II
pASC1485	HI0936	cytochrome C-type biogenesis	995057	362	635 *Pvu*II
pASC1486	HI0147	conserved hypothetical transmembrane protein	163459	300	633 *Pvu*II
pASC1487	HI0936	cytochrome C-type biogenesis	995057	362	635 *Pvu*II

Gray shading indicates multiple insertions mapped in the same sequencing reaction

Appendix C – Mapped Tn5 insertions (pASC15 libraries)

Allele/Plasmid	CDS/Vector	Protein/Gene	Mutation coordinate	Disrupted codon	Total codons	Percent GC	Distance from junct (<300nt) †	Library	Comments
pASC1220	pASC15	vector	2259					*Pvu*II	
pASC1221	HI0380	conserved hypothetical protein	401360	108	241	35.13	*117*	*Pvu*II	
pASC1222	HI1012	sugar isomerase, putative	1076700	40	210	40.16		*Pvu*II	
pASC1223	HI1662	2-oxoglutarate dehydrogenase E1 component (sucA)	1731673	227	950	41.09		*Pvu*II	
pASC1224	pASC15	vector	181					*Pvu*II	
pASC1226	HI0594	conserved hypothetical transmembrane protein	616080	407	509	39.75		*Pvu*II	
pASC1227	HI0832	fumarate reductase, 13 kDa hydrophobic protein (frdD)	882437	84	114	39.18	**19**	*Pvu*II	
pASC1228	pASC15	vector	1227					*Pvu*II	
pASC1229	pASC15	vector	267					*Pvu*II	
pASC1230	HI1011	conserved hypothetical protein	1075575	78	413	41		*Pvu*II	
pASC1231	pASC15	vector	480					*Pvu*II	
pASC1232	pASC15	vector	416					*Pvu*II	
pASC1233	pASC15	vector	526					*Pvu*II	
pASC1234	pASC15	vector	558					*Pvu*II	
pASC1235	HI1501	H. influenzae predicted coding region HI1501	1575513	483	520	48.33		*Pvu*II	
pASC1236	HI0148	conserved hypothetical protein	165705	360	379	40.28		*Pvu*II	
pASC1237	pASC15	vector	304					*Pvu*II	
pASC1239	rrn	rRNA operon	rrn					*Pvu*II	
pASC1240	pASC15	vector	200					*Pvu*II	
pASC1241	pASC15	vector	181					*Pvu*II	
pASC1242	HI0898	multidrug resistance protein A (emrA)	954668	227	390	38.72		*Pvu*II	
pASC1243	HI1054	conserved hypothetical protein	1119067	446	460	36.09		*Pvu*II	
pASC1244	HI1501	H. influenzae predicted coding region HI1501	1575624	520	520	48.33		*Pvu*II	
pASC1245	HI1500	H. influenzae predicted coding region HI1500	1573706	394	508	47.9		*Pvu*II	
pASC1246	pASC15	vector	198					*Pvu*II	

pASC1247	pASC15	vector	2269				*Pvu*II	
pASC1248	pASC15	vector	579				*Pvu*II	
pASC1249	HI0947	virulence associated protein C (vapC)	1008468	38	132	37.12	*Pvu*II	
pASC1250	pASC15	vector	430				*Pvu*II	
pASC1251	pASC15	vector	551				*Pvu*II	
pASC1252	HI0617	conserved hypothetical protein	651670	191	219	42.31	*193* *Pvu*II	
pASC1253	HI1059	ribonuclease HII (rnhB)	1124613	81	197	42.64	*Pvu*II	
pASC1254	pASC15	vector	186				*Pvu*II	
pASC1255	HI0203	conserved hypothetical protein	218679	17	178	36.89	*Pvu*II	
pASC1256	pASC15	vector	1227				*Pvu*II	
pASC1257	rrn	rRNA operon	rrn				*Pvu*II	
pASC1258	HI0594	conserved hypothetical transmembrane protein	615643	129	509	39.75	*Pvu*II	Intergenic
pASC1258	HI0594	conserved hypothetical transmembrane protein	615939	454	509	39.75	*Pvu*II	
pASC1260	pASC15	vector	237				*Pvu*II	
pASC1262	HI1186	dipeptide ABC transporter, permease protein (dppC)	1251746	290	295	41.47	*Pvu*II	
pASC1263	HI1186	dipeptide ABC transporter, permease protein (dppC)	1252342	91	295	41.47	*Pvu*II	
pASC1264	HI0380	conserved hypothetical protein	401416	89	241	35.13	**123** *Pvu*II	
pASC1265	pASC15	vector	236				*Pvu*II	
pASC1266	pASC15	vector	359				*Pvu*II	
pASC1267	pASC15	vector	281				*Pvu*II	
pASC1268	HI1662	2-oxoglutarate dehydrogenase E1 component (sucA)	1731266	363	950	41.09	**41** *Pvu*II	
pASC1269	HI1012	sugar isomerase, putative	1076683	34	210	40.16	*Pvu*II	
pASC1270	pASC15	vector	38				*Pvu*II	
pASC1271	pASC15	vector	470				*Pvu*II	*Hinf* junct. @ 614365
pASC1272	HI1107	Na+/H+ antiporter (nhaC)	1172149	128	468	35.97	*157* *Pvu*II	
pASC1273	HI0834	fumarate reductase, iron-sulfur protein (frdB)	883447	149	256	39.45	*Pvu*II	
pASC1274	HI1501	H. influenzae predicted coding region HI1501	1575041	326	520	48.33	*Pvu*II	
pASC1275	HI1056	H. influenzae predicted coding region HI1056	1122302	462	629	35.19	*Pvu*II	
pASC1276	pASC15	vector	2171				*Pvu*II	
pASC1277	HI0380	conserved hypothetical protein	401279	135	241	35.13	**36** *Pvu*II	
pASC1278	HI0594	conserved hypothetical transmembrane protein	616987	104	509	39.75	*Pvu*II	
pASC1279	pASC15	vector	1662				*Pvu*II	

pASC1280	HI0898	multidrug resistance protein A (emrA)	954728	207	390	38.72	*188* *Pvu*II	
pASC1282	HI0380	conserved hypothetical protein	401263	140	241	35.13	**20** *Pvu*II	
pASC1283	rrn	rRNA operon	rrn				*Pvu*II	
pASC1284	pASC15	vector	13				*Pvu*II	
pASC1285	HI0617	conserved hypothetical protein	651176	26	219	42.31	*Pvu*II	
pASC1286	HI0166	NADH:ubiquinone oxidoreductase, subunit B (nqrB)	181550	201	411		*Pvu*II	
pASC1287	pASC15	vector	134				*Pvu*II	
pASC1288	HI0346	ferredoxin-type protein (napH)	374097	97	287	41.93	*Pvu*II	
pASC1289	HI1662	2-oxoglutarate dehydrogenase E1 component (sucA)	1731737	206	950	41.09	*Pvu*II	
pASC1290	pASC15	vector	556				*Pvu*II	
pASC1291	pASC15	vector	189				*Pvu*II	
pASC1292	pASC15	vector	1250				*Pvu*II	
pASC1292	HI0946	formamidopyrimidine-DNA glycosylase (fpg)	1005814	194	271	37.88	*Pvu*II	
pASC1292	HI0946	formamidopyrimidine-DNA glycosylase (fpg)	1005821	192	271	37.88	*Pvu*II	
pASC1293	HI0171	Nqr6 sub of NADH-quinone red complex beta-sub (nqr6)	184771	208	411	39.58	*Pvu*II	
pASC1294	rrn	rRNA operon	rrn				*Pvu*II	
pASC1295	pASC15	vector	352				*Pvu*II	*Hinf* junct. @ 1573195
pASC1296	pASC15	vector	353				*Pvu*II	
pASC1297	HI1011	conserved hypothetical protein	1076013	224	413	41	*Pvu*II	
pASC1297	HI1012	sugar isomerase, putative	1076703	41	210	40.16	*Pvu*II	
pASC1298	rrn	rRNA operon	rrn				*Pvu*II	
pASC1299	pASC15	vector	1222				*Pvu*II	
pASC1301	HI0946	formamidopyrimidine-DNA glycosylase (fpg)	1006048	116	271	37.88	*Pvu*II	
pASC1302	rrn	rRNA operon	rrn				*Pvu*II	
pASC1303	pASC15	vector	47				*Pvu*II	
pASC1304	pASC15	vector	551				*Pvu*II	
pASC1305	HI0147	conserved hypothetical transmembrane protein	163470	304	633	37.86	*Pvu*II	
pASC1306	pASC15	vector	2293				*Pvu*II	
pASC1307	pASC15	vector	2259				*Pvu*II	
pASC1308	pASC15	vector	2292				*Pvu*II	
pASC1309	HI1013	conserved hypothetical protein	1077365	50	258	38.76	**36** *Pvu*II	
pASC1310	rrn	rRNA operon	rrn				*Pvu*II	

pASC1311	pASC15	vector	308				*Pvu*II	
pASC1312	pASC15	vector	139				*Pvu*II	
pASC1313	pASC15	vector	2261				*Pvu*II	
pASC1314	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1006921	414	511	42.14	*61 Pvu*II	
pASC1315	HI0833	fumarate reductase, 15 kDa hydrophobic protein (frdC)	882973	42	132	36.36	*Pvu*II	
pASC1316	pASC15	vector	136				*Pvu*II	
pASC1317	HI1501	H. influenzae predicted coding region HI1501	1574289	75	520	48.33	*Pvu*II	
pASC1317	HI1502	H. influenzae predicted coding region HI1502	1576372	219	414	46.38	*Pvu*II	
pASC1318	HI0591	ornithine decarboxylase (speF)	614521	88	720	36.85	*155 Pvu*II	
pASC1319	HI1500	H. influenzae predicted coding region HI1500	1573451	309	508	47.9	*276 Pvu*II	
pASC1320	pASC15	vector	198				*Pvu*II	
pASC1321	pASC15	vector	359				*Pvu*II	
pASC1322	HI0558	glucose-6-phosphate 1-dehydrogenase (zwf)	577416	319	494	40.96	*Pvu*II	
pASC1323	HI0898	multidrug resistance protein A (emrA)	954656	231	390	38.72	*Pvu*II	
pASC1501	HI0896	cell division protein (ftsN)	952610	202	204	37.91	*Pvu*II	
pASC1502	HI1011	conserved hypothetical protein	1076116	258	413	41	*Pvu*II	
pASC1503	HI0897	multidrug resistance protein B (emrB)	953812	118	510	39.67	*210 Pvu*II	
pASC1504	HI1011	conserved hypothetical protein	1075775	144	413	41	*Pvu*II	
pASC1505	HI0834	fumarate reductase, iron-sulfur protein (frdB)	883556	112	256	39.45	*Pvu*II	
pASC1507	HI0946	formamidopyrimidine-DNA glycosylase (fpg)	1006153	81	271	37.88	*Pvu*II	
pASC1508	HI0595	carbamate kinase (arcC)	617731	202	310	40.54	*Pvu*II	
pASC1509	pASC15	vector	2118				*Pvu*II	
pASC1510	HI0895	acriflavine resistance protein (acrB)	951021	835	1032	38.63	*Pvu*II	
pASC1512	rrn	rRNA operon	rrn				*Pvu*II	
pASC1513	HI1502	H. influenzae predicted coding region HI1502	1575632	83	414	46.38	*Pvu*II	Intergenic
pASC1514	HI0380	conserved hypothetical protein	401538	49	241	35.13	*1 Pvu*II	
pASC1516	HI1010	3-hydroxyisobutyrate dehydrogenase, putative	1075168	245	301	40.31	*89 Pvu*II	
pASC1517	HI0594	conserved hypothetical transmembrane protein	615602	170	509	39.75	*Pvu*II	Intergenic
pASC1518	HI1059	ribonuclease HII (rnhB)	1124119	145	197	42.64	*Pvu*II	Intergenic
pASC1519	pASC15	vector	1214				*Pvu*II	
pASC1520	HI1185	dipeptide ABC transporter, ATP-binding protein (dppD)	1251645	24	330	42.73	*Pvu*II	
pASC1522	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1008087	26	511	42.14	*Pvu*II	

pASC1523	HI0948	conserved hypothetical protein	1008585	77	77	39.83	*Pvu*II
pASC1524	HI0946	formamidopyrimidine-DNA glycosylase (fpg)	1006038	119	271	37.88	*Pvu*II
pASC1525	pASC15	vector	307				*Pvu*II
pASC1526	HI0617	conserved hypothetical protein	651408	104	219	42.31	*Pvu*II
pASC1527	HI1502	H. influenzae predicted coding region HI1502	1576369	218	414	46.38	*Pvu*II
pASC1528	HI0832	fumarate reductase, 13 kDa hydrophobic protein (frdD)	882625	21	114	39.18	*205* *Pvu*II
pASC1529	HI1502	H. influenzae predicted coding region HI1502	1576621	302	414	46.38	**82** *Pvu*II
pASC1530	pASC15	vector	307				*Pvu*II
pASC1531	HI0949	aminotransferase	1009611	238	454	43.1	**59** *Pvu*II
pASC1532	HI1662	2-oxoglutarate dehydrogenase E1 component (sucA)	1731428	309	950	41.09	*210* *Pvu*II
pASC1533	HI0897	multidrug resistance protein B (emrB)	953957	70	510	39.67	*Pvu*II
pASC1534	HI0898	multidrug resistance protein A (emrA)	954294	351	390	38.72	*Pvu*II
pASC1535	HI1501	H. influenzae predicted coding region HI1501	1574550	162	520	48.33	*Pvu*II
pASC1536	rrn	rRNA operon	rrn				*Pvu*II
pASC1537	HI1053	conserved hypothetical protein	1118871	96	113	43.07	*179* *Pvu*II
pASC1538	HI1185	dipeptide ABC transporter, ATP-binding protein (dppD)	1251662	19	330	42.73	*Pvu*II
pASC1539	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007872	97	511	42.14	*Pvu*II
pASC1541	HI0591	ornithine decarboxylase (speF)	614699	29	720	36.85	*Pvu*II
pASC1542	HI1011	conserved hypothetical protein	1076013	224	413	41	*Pvu*II
pASC1543	HI0594	conserved hypothetical transmembrane protein	616116	395	509	39.75	*Pvu*II
pASC1544	HI1662	2-oxoglutarate dehydrogenase E1 component (sucA)	1731980	125	950	41.09	**77** *Pvu*II
pASC1545	HI1186	dipeptide ABC transporter, permease protein (dppC)	1251980	212	295	41.47	*Pvu*II
pASC1546	pASC15	vector	1262				*Pvu*II
pASC1547	pASC15	vector	1660				*Pvu*II
pASC1548	pASC15	vector	307				*Pvu*II
pASC1601	pASC15	vector	307				*Pvu*II
pASC1602	pASC15	vector	307				*Pvu*II
pASC1603	HI0946	formamidopyrimidine-DNA glycosylase (fpg)	1006045	117	271	37.88	*Pvu*II
pASC1604	pASC15	vector	307				*Pvu*II
pASC1605	pASC15	vector	307				*Pvu*II
pASC1606	pASC15	vector	307				*Pvu*II
pASC1608	pASC15	vector	307				*Pvu*II

pASC1609	HI0946	formamidopyrimidine-DNA glycosylase (fpg)	1006045	117	271	37.88	*Pvu*II
pASC1610	pASC15	vector	307				*Pvu*II
pASC1611	HI0946	formamidopyrimidine-DNA glycosylase (fpg)	1006045	117	271	37.88	*Pvu*II
pASC1612	pASC15	vector	307				*Pvu*II
pASC1613	pASC15	vector	307				*Pvu*II
pASC1614	pASC15	vector	307				*Pvu*II
pASC1617	pASC15	vector	307				*Pvu*II
pASC1619	pASC15	vector	307				*Pvu*II
pASC1620	pASC15	vector	307				*Pvu*II
pASC1621	pASC15	vector	307				*Pvu*II
pASC1622	pASC15	vector	307				*Pvu*II
pASC1623	pASC15	vector	307				*Pvu*II
pASC1624	pASC15	vector	307				*Pvu*II

Gray shading indicates multiple insertions mapped in the same sequencing reaction

[†] Distance of Tn5 insertion from insert/vector junction (**bold** = experimental; *italics* = theoretical)

Appendix D – Mapped Tn5 insertions (pASC18, pre-mutagenesis minimalization)

Allele/Plasmid	CDS/Vector	Protein/Gene	Mutation coordinate	disrupted codon	Total codons	Percent GC	Distance from junct (<300nt) †	Library	Comments
pASC1701	pASC18	vector	483					*Eco*RI	
pASC1702	pASC18	vector	1400					*Eco*RI	
pASC1703	pASC18	vector	401					*Eco*RI	
pASC1704	pASC18	vector	395					*Eco*RI	
pASC1705	pASC18	vector	395					*Eco*RI	
pASC1706	pASC18	vector	401					*Eco*RI	
pASC1707	pASC18	vector	395					*Eco*RI	
pASC1708	HI0602	hemY protein (hemY)	630612	340	428	36.84		*Eco*RI	
pASC1709	pASC18	vector	401					*Eco*RI	
pASC1710	pASC18	vector	483					*Eco*RI	
pASC1713	pASC18	vector	401					*Eco*RI	
pASC1714	pASC18	vector	395					*Eco*RI	
pASC1715	HI0357	thiamine biosynthesis protein, putative	382122	0	314	39.7		*Eco*RI	
pASC1717	pASC18	vector	1286					*Eco*RI	
pASC1718	HI1447	GTP cyclohydrolase I (folE)	1534184	148	218	36.7		*Eco*RI	
pASC1719	HI0963	riboflavin kinase / FMN adenylyltransferase (ribF)	1022089	183	312	39.21	*286*	*Eco*RI	
pASC1720	rrn	rRNA operon	rrn					*Eco*RI	
pASC1721	HI1548	conserved hypothetical transmembrane protein	1619890	176	416	38.06		*Eco*RI	
pASC1722	pASC18	vector	401					*Eco*RI	
pASC1723	pASC18	vector	998					*Eco*RI	
pASC1724	pASC18	vector	401					*Eco*RI	
pASC1725	pASC18	vector	401					*Eco*RI	
pASC1726	pASC18	vector	394					*Eco*RI	
pASC1727	HI0729	prolyl-tRNA-synthetase (proS)	782638	24	572	43.12		*Eco*RI	
pASC1729	pASC18	vector	401					*Eco*RI	
pASC1730	pASC18	vector	401					*Eco*RI	
pASC1732	pASC18	vector	393					*Eco*RI	
pASC1734	pASC18	vector	359					*Eco*RI	

pASC1736	HI0691	glycerol kinase (glpK)	735473	129	503	42.68		*Eco*RI	
pASC1738	pASC18	vector	402					*Eco*RI	
pASC1739	pASC18	vector	404					*Eco*RI	
pASC1749	rrn	rRNA operon	rrn					*Pvu*II	
pASC1750	HI1702	5-methyltetrahydropteroyltriglutamate-homocysteine methyltransferase (metE)	1772505	530	756	40.08		*Pvu*II	
pASC1751	HI0231	ATP-dependent RNA helicase (deaD)	261730	52	613	42.03		*Pvu*II	
pASC1752	HI0006	formate dehydrogenase, alpha subunit (fdxG)	6573	305	1029			*Pvu*II	
pASC1753	HI0006	formate dehydrogenase, alpha subunit (fdxG)	6840	394	1029			*Pvu*II	
pASC1754	HI0932	enolase (eno)	990080	384	436	41.74		*Pvu*II	
pASC1755	HI0358	transcriptional activator, putative	383721	215	215	42.33		*Pvu*II	
pASC1756	HI1502	H. influenzae predicted coding region HI1502	1576747	344	414	46.38	**46**	*Pvu*II	
pASC1757	HI1522	H. influenzae predicted coding region HI1522	1591022	40	623	41.63		*Pvu*II	
pASC1758	HI1502	H. influenzae predicted coding region HI1502	1576901	395	414	46.38	*198*	*Pvu*II	
pASC1759	HI0748	glycerol-3-phosphate acyltransferase (plsB)	808281	515	810	36.01		*Pvu*II	
pASC1760	pASC18	vector	1277					*Pvu*II	
pASC1762	HI0928	catalase (hktE)	987131	414	508	42.72		*Pvu*II	
pASC1763	HI1011	conserved hypothetical protein	1075439	32	413	41		*Pvu*II	
pASC1764	HI0359	iron (chelated) ABC transporter, permease protein (yfeD)	384030	202	271	35.79		*Pvu*II	
pASC1765	HI1022	biotin synthetase (bioB)	1086700	188	333	41.74		*Pvu*II	
pASC1766	HI0614	L-fucose isomerase (fucI)	645953	413	604	41.17	**13**	*Pvu*II	
pASC1767	HI1160	ferrochelatase (hemH)	1229407	132	323	38.08		*Pvu*II	Intergenic
pASC1768	HI0579	elongation factor G (fusA)	597931	639	700	40.67	*156*	*Pvu*II	
pASC1768	HI1054	conserved hypothetical protein	1119486	306	460	36.09		*Pvu*II	
pASC1769	rrn	rRNA operon	rrn					*Pvu*II	
pASC1770	HI0413	ribonuclease E (rne)	434434	99	951	40.41		*Pvu*II	
pASC1771	HI1501	H. influenzae predicted coding region HI1501	1574138	25	520	48.33		*Pvu*II	
pASC1772	HI0623	methionyl-tRNA-formyltransferase (fmt)	663199	12	318	38.78	*130*	*Pvu*II	Intergenic
pASC1773	HI0448	PTS system, fructose-specific IIA/FPr component (fruB)	472182	65	499	41.88		*Pvu*II	
pASC1774	HI1023	transketolase 1 (tktA)	1087609	550	665	43.06		*Pvu*II	
pASC1774	HI1023	transketolase 1 (tktA)	1087952	435	665	43.06		*Pvu*II	
pASC1775	HI0712	hemoglobin-binding protein	760815	173	1084	36.19		*Pvu*II	
pASC1776	HI0579	elongation factor G (fusA)	598156	564	700	40.67		*Pvu*II	
pASC1777	HI0928	catalase (hktE)	987229	447	508	42.72		*Pvu*II	

pASC1778	HI1160	ferrochelatase (hemH)	1230015	165	323	38.08		*Pvu*II	
pASC1779	pASC18	vector	473					*Pvu*II	Hinf junct @ 629536
pASC1780	HI0713	trigger factor (tig)	762698	353	432	36.42		*Pvu*II	
pASC1781	pASC18	vector	1886					*Pvu*II	
pASC1782	HI0928	catalase (hktE)	985989	33	508	42.72	*7*	*Pvu*II	
pASC1783	HI1522	H. influenzae predicted coding region HI1522	1590950	16	623	41.63		*Pvu*II	
pASC1784	rrn	rRNA operon	rrn					*Pvu*II	
pASC1785	HI0156	malonyl CoA-acyl carrier protein transacylase (fabD)	172543	185	312	41.99		*Pvu*II	
pASC1786	HI1516	phage related protein frameshift mutation	1587292	104	419			*Pvu*II	
pASC1787	HI0622	polypeptide deformylase (def)	663004	131	169	37.08		*Pvu*II	
pASC1787	rrn	rRNA operon	rrn					*Pvu*II	
pASC1788	HI0897	multidrug resistance protein B (emrB)	952843	441	510	39.67		*Pvu*II	
pASC1789	pASC18	vector	1509					*Pvu*II	
pASC1790	HI0410	transcriptional regulatory protein (tyrR)	431648	90	318	37.53		*Pvu*II	
pASC1791	HI0153	dcuB frameshift	170255	285	444		**24**	*Pvu*II	
pASC1792	HI1365	DNA topoisomerase I (topA)	1452442	501	868	39.21		*Pvu*II	
pASC1793	HI0579	elongation factor G (fusA)	598149	566	700	40.67		*Pvu*II	
pASC1794	HI1018	IS1016-V6 protein frameshift mutation	1082190	108	173		**54**	*Pvu*II	
pASC1795	pASC18	vector	1382					*Pvu*II	
pASC1795	pASC18	vector	2135					*Pvu*II	
pASC1796	HI1354	glutaminyl-tRNA-synthetase (glnS)	1433087	226	557	39.5		*Pvu*II	
pASC1797	HI0229	polynucleotide phosphorylase (pnp)	259400	368	709	42.5	*260*	*Pvu*II	
pASC1798	HI0579	elongation factor G (fusA)	598168	560	700	40.67		*Pvu*II	
pASC1799	HI0579	elongation factor G (fusA)	597840	669	700	40.67	*65*	*Pvu*II	
pASC1799	HI0579	elongation factor G (fusA)	598363	495	700	40.67		*Pvu*II	
pASC1800	HI0579	elongation factor G (fusA)	598131	572	700	40.67		*Pvu*II	
pASC1801	HI0672	conserved hypothetical protein	716181	36	225	40.44	**79**	*Pvu*II	
pASC1803	pASC18	vector	1396					*Pvu*II	
pASC1804	HI0928	catalase (hktE)	987016	376	508	42.72		*Pvu*II	
pASC1805	HI0448	PTS system, fructose-specific IIA/FPr component (fruB)	470886	497	499	41.88	*206*	*Pvu*II	
pASC1806	HI0928	catalase (hktE)	987057	389	508	42.72		*Pvu*II	
pASC1806	HI0929	conserved hypothetical protein	988507	106	393	37.49		*Pvu*II	
pASC1807	HI0897	multidrug resistance protein B (emrB)	952843	441	510	39.67		*Pvu*II	
pASC1808	HI0579	elongation factor G (fusA)	598168	560	700	40.67		*Pvu*II	
pASC1809	rrn	rRNA operon	rrn				**57**	*Pvu*II	

pASC1810	pASC18	vector	756					*Pvu*II	
pASC1810	pASC18	vector	1518					*Pvu*II	
pASC1810	pASC18	vector	1900					*Pvu*II	
pASC1811	HI0928	catalase (hktE)	987400	504	508	42.72		*Pvu*II	
pASC1812	HI0579	elongation factor G (fusA)	597944	634	700	40.67	*169*	*Pvu*II	
pASC1813	HI1516	phage related protein frameshift mutation	1587484	168	419		*116*	*Pvu*II	
pASC1814	pASC18	vector	2268					*Pvu*II	
pASC1815	pASC18	vector	1293					*Pvu*II	
pASC1816	rrn	rRNA operon	rrn				**42**	*Pvu*II	
pASC1817	HI0559.1	H. influenzae predicted coding region HI0559.1	579193	139	115	39.42		*Pvu*II	Intergenic
pASC1817	HI0559.1	H. influenzae predicted coding region HI0559.1	579329	3	115	39.42		*Pvu*II	Intergenic
pASC1818	HI1518	H. influenzae predicted coding region HI1518	1588313	12	182	50.18	*200*	*Pvu*II	
pASC1819	HI0622	polypeptide deformylase (def)	662870	87	169	37.08		*Pvu*II	
pASC1819	HI0622	polypeptide deformylase (def)	662877	89	169	37.08		*Pvu*II	
pASC1820	HI0579	elongation factor G (fusA)	598392	485	700	40.67		*Pvu*II	
pASC1821	rrn	rRNA operon	rrn					*Pvu*II	
pASC1822	rrn	rRNA operon	rrn				**59**	*Pvu*II	
pASC1823	pASC18	vector	2133					*Pvu*II	
pASC1824	pASC18	vector	1358					*Pvu*II	
pASC1825	HI1502	H. influenzae predicted coding region HI1502	1575660	55	414	46.38		*Pvu*II	Intergenic
pASC1826	HI0928	catalase (hktE)	987395	502	508	42.72		*Pvu*II	
pASC1827	pASC18	vector	1265					*Pvu*II	
pASC1828	HI0579	elongation factor G (fusA)	598291	519	700	40.67		*Pvu*II	
pASC1829	rrn	rRNA operon	rrn					*Pvu*II	
pASC1830	HI0614	L-fucose isomerase (fucI)	645925	422	604	41.17	**41**	*Pvu*II	
pASC1831	pASC18	vector	1293					*Pvu*II	
pASC1832	HI0579	elongation factor G (fusA)	598293	518	700	40.67		*Pvu*II	
pASC1833	HI0930	H. influenzae predicted coding region HI0930	988932	172	206	41.59	*228*	*Pvu*II	
pASC1833	HI0930	H. influenzae predicted coding region HI0930	988939	169	206	41.59	*221*	*Pvu*II	
pASC1834	rrn	rRNA operon	rrn					*Pvu*II	
pASC1835	HI0929	conserved hypothetical protein	987415	231	393	37.49		*Pvu*II	Intergenic
pASC1835	HI0930	H. influenzae predicted coding region HI0930	988886	187	206	41.59	*274*	*Pvu*II	
pASC1836	HI0931	H. influenzae predicted coding region HI0931	989564	114	161	38.3		*Pvu*II	
pASC1836	HI0932	enolase (eno)	990394	280	436	41.74		*Pvu*II	
pASC1837	HI0929	conserved hypothetical protein	988786	13	393	37.49		*Pvu*II	
pASC1839	HI0928	catalase (hktE)	987023	378	508	42.72		*Pvu*II	

pASC1840	HI0930	H. influenzae predicted coding region HI0930	989138	103	206	41.59	*22*	*Pvu*II	
pASC1841	HI0929	conserved hypothetical protein	987534	112	393	37.49		*Pvu*II	Intergenic
pASC1841	HI1504	I protein (muI)	1578776	336	355	47.79		*Pvu*II	
pASC1842	pASC18	vector	1774					*Pvu*II	
pASC1843	HI1520	conserved hypothetical protein	1589827	195	355	48.26	*18*	*Pvu*II	
pASC1844	HI1520	conserved hypothetical protein	1589834	197	355	48.26	**27**	*Pvu*II	
pASC1845	HI0672	conserved hypothetical protein	715963	109	225	40.44	*297*	*Pvu*II	
pASC1846	rrn	rRNA operon	rrn					*Pvu*II	
pASC1848	HI0579	elongation factor G (fusA)	597944	634	700	40.67	*169*	*Pvu*II	
pASC1849	HI1520	conserved hypothetical protein	1589835	198	355	48.26	**28**	*Pvu*II	
pASC1850	pASC18	vector	1521					*Pvu*II	
pASC1850	pASC18	vector	2133					*Pvu*II	
pASC1851	rrn	rRNA operon	rrn					*Pvu*II	
pASC1852	HI0361	iron (chelated) transporter, ATP-binding protein (yfeB)	385738	221	306	39.43		*Pvu*II	
pASC1852	HI0613	fuculokinase (fucK)	644423	294	470	38.51		*Pvu*II	
pASC1853	HI0229	polynucleotide phosphorylase (pnp)	260163	622	709	42.5		*Pvu*II	
pASC1854	HI0932	enolase (eno)	990341	297	436	41.74		*Pvu*II	
pASC1855	HI1501	H. influenzae predicted coding region HI1501	1574497	145	520	48.33		*Pvu*II	
pASC1856	rrn	rRNA operon	rrn					*Pvu*II	
pASC1859	HI1502	H. influenzae predicted coding region HI1502	1575826	37	414	46.38		*Pvu*II	
pASC1860	rrn	rRNA operon	rrn					*Pvu*II	
pASC1861	HI0448	PTS system, fructose-specific IIA/FPr component (fruB)	471951	142	499	41.88		*Pvu*II	
pASC1861	HI1520	conserved hypothetical protein	1590059	272	355	48.26	*250*	*Pvu*II	
pASC1862	HI1501	H. influenzae predicted coding region HI1501	1574182	40	520	48.33		*Pvu*II	
pASC1863	HI0579	elongation factor G (fusA)	598470	459	700	40.67	*235*	*Pvu*II	
pASC1864	HI0037	rod shape-determining protein (mreB)	40314	81	378	40.04	**8**	*Pvu*II	
pASC1865	HI0344	napA protein frameshift	372022	536	832			*Pvu*II	
pASC1866	rrn	rRNA operon	rrn					*Pvu*II	
pASC1867	pASC18	vector	263					*Pvu*II	
pASC1868	pASC18	vector	254					*Pvu*II	
pASC1869	pASC18	vector	1293					*Pvu*II	
pASC1870	HI1214	single-stranded-DNA-specific exonuclease (recJ)	1282208	270	575	40.35		*Pvu*II	
pASC1871	pASC18	vector	402					*Pvu*II	
pASC1872	HI0037	rod shape-determining protein (mreB)	40308	81	378	40.04	**8**	*Pvu*II	

pASC1873	HI0947	virulence associated protein C (vapC)	1008269	104	132	37.12		*Pvu*II	
pASC1874	HI1161	conserved hypothetical protein	1230739	61	138	42.51		*Pvu*II	
pASC1876	pASC18	vector	1320					*Pvu*II	
pASC1877	rrn	rRNA operon	rrn					*Pvu*II	
pASC1878	HI1520	conserved hypothetical protein	1589575	111	355	48.26	*234*	*Pvu*II	
pASC1879	HI0748	glycerol-3-phosphate acyltransferase (plsB)	807381	13	810	36.01		*Pvu*II	Intergenic
pASC1880	HI0928	catalase (hktE)	987098	403	508	42.72		*Pvu*II	
pASC1881	HI1520	conserved hypothetical protein	1589827	195	355	48.26	*18*	*Pvu*II	
pASC1882	HI0411	host factor-I protein (hfq)	432688	5	91	37.73		*Pvu*II	
pASC1883	pASC18	vector	427					*Pvu*II	Hinf junct @ 985982
pASC1885	HI0411	host factor-I protein (hfq)	432546	53	91	37.73		*Pvu*II	
pASC1886	HI0928	catalase (hktE)	986001	37	508	42.72	*19*	*Pvu*II	
pASC1888	HI1163	conserved hypothetical protein	1232575	785	1027	39.21	*66*	*Pvu*II	
pASC1889	HI0222	GMP synthase (guaA)	252293	389	523	43.66	**9**	*Pvu*II	
pASC1890	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1006900	421	511	42.14	**42**	*Pvu*II	
pASC1891	pASC18	vector	1621					*Pvu*II	
pASC1891	pASC18	vector	1628					*Pvu*II	
pASC1892	pASC18	vector	1265					*Pvu*II	
pASC1893	HI0231	ATP-dependent RNA helicase (deaD)	262764	397	613	42.03	*189*	*Pvu*II	
pASC1894	HI0928	catalase (hktE)	986091	67	508	42.72	*109*	*Pvu*II	
pASC1894	HI0928	catalase (hktE)	986563	225	508	42.72		*Pvu*II	
pASC1895	HI0440	penicillin-binding protein 1A (ponA)	462068	474	864	41.32		*Pvu*II	
pASC1895	HI0440	penicillin-binding protein 1A (ponA)	462531	628	864	41.32	*252*	*Pvu*II	
pASC1896	pASC18	vector	572					*Pvu*II	
pASC1896	HI0669	mioC protein (mioC)	714000	11	146	39.04	*198*	*Pvu*II	
pASC1897	HI0928	catalase (hktE)	987379	497	508	42.72		*Pvu*II	
pASC1898	HI0140	N-acetylglucosamine-6-phosphate deacetylase (nagA)	155414	346	381	38.58	*216*	*Pvu*II	
pASC1899	rrn	rRNA operon	rrn					*Pvu*II	
pASC1900	pASC18	vector	1293					*Pvu*II	
pASC1901	HI0037	rod shape-determining protein (mreB)	40314	81	378	40.04	**8**	*Pvu*II	
pASC1902	HI1521	conserved hypothetical protein	1590311	83	168	41.07		*Pvu*II	Intergenic
pASC1903	rrn	rRNA operon	rrn					*Pvu*II	
pASC1904	HI0449	H. influenzae predicted coding region HI0449	472798	63	178	37.27	*33*	*Pvu*II	
pASC1905	pASC18	vector	2278					*Pvu*II	
pASC1906	HI1023	transketolase 1 (tktA)	1087539	573	665	43.06		*Pvu*II	
pASC1907	HI1504	I protein (muI)	1577966	66	355	47.79		*Pvu*II	

pASC1908	rrn	rRNA operon	rrn					*Pvu*II
pASC1909	HI1520	conserved hypothetical protein	1590051	270	355	48.26	*242*	*Pvu*II
pASC1910	pASC18	vector	1452					*Pvu*II
pASC1910	pASC18	vector	1607					*Pvu*II
pASC1911	HI1502	H. influenzae predicted coding region HI1502	1576492	259	414	46.38	*211*	*Pvu*II
pASC1912	HI0412	conserved hypothetical protein	433405	119	322	38.61		*Pvu*II
pASC1912	HI0413	ribonuclease E (rne)	434337	66	951	40.41		*Pvu*II
pASC1913	HI0358	transcriptional activator, putative	383632	185	215	42.33		*Pvu*II
pASC1914	pASC18	vector	1383					*Pvu*II
pASC1915	HI0819	galactokinase (galK)	870993	38	397	40.97	**9**	*Pvu*II
pASC1915	HI1501	H. influenzae predicted coding region HI1501	1574182	40	520	48.33		*Pvu*II
pASC1916	pASC18	vector	1448					*Pvu*II
pASC1917	HI0411	host factor-I protein (hfq)	432638	22	91	37.73		*Pvu*II
pASC1918	HI1521	conserved hypothetical protein	1590764	123	168	41.07		*Pvu*II
pASC1919	HI0361	iron (chelated) transporter, ATP-binding protein (yfeB)	386336	22	306	39.43		*Pvu*II
pASC1919	HI0558	glucose-6-phosphate 1-dehydrogenase (zwf)	577114	420	494	40.96	*152*	*Pvu*II
pASC1919	HI0561	conserved hypothetical protein	580028	536	633	41.13		*Pvu*II
pASC1920	rrn	rRNA operon	rrn					*Pvu*II
pASC1921	HI0229	polynucleotide phosphorylase (pnp)	259528	410	709	42.5		*Pvu*II
pASC1922	HI1520	conserved hypothetical protein	1589965	241	355	48.26	*24*	*Pvu*II
pASC1923	HI1520	conserved hypothetical protein	1589270	9	355	48.26		*Pvu*II
pASC1924	pASC18	vector	427					*Pvu*II
pASC1925	HI1159	conserved hypothetical protein	1228996	86	217	37.33	*186*	*Pvu*II
pASC1926	HI1520	conserved hypothetical protein	1589263	7	355	48.26		*Pvu*II
pASC1927	HI1023	transketolase 1 (tktA)	1087793	488	665	43.06		*Pvu*II
pASC1928	pASC18	vector	1021					*Pvu*II
pASC1928	HI0602	hemY protein (hemY)	630707	308	428	36.84		*Pvu*II
pASC1928	HI0603	uroporphyrin-III C-methyltransferase (hemX)	632562	98	230	36.52	*237*	*Pvu*II
pASC1929	HI1053	conserved hypothetical protein	1118889	102	113	43.07	*197*	*Pvu*II
pASC1930	rrn	rRNA operon	rrn					*Pvu*II
pASC1932	HI0140	N-acetylglucosamine-6-phosphate deacetylase (nagA)	155534	306	381	38.58	**96**	*Pvu*II
pASC1933	HI0579	elongation factor G (fusA)	597833	671	700	40.67	**60**	*Pvu*II
pASC1937	pASC18	vector	433					*Fsp*I
pASC1938	pASC18	vector	433					*Fsp*I
pASC1939	pASC18	vector	427					*Fsp*I

pASC1940	pASC18	vector	427					*Fsp*I
pASC1941	pASC18	vector	433					*Fsp*I
pASC1942	pASC18	vector	433					*Fsp*I
pASC1943	pASC18	vector	427					*Fsp*I
pASC1944	pASC18	vector	427					*Fsp*I
pASC1945	HI1735	peptide chain release factor 3 (prfC)	1812966	240	527	41.68		*Fsp*I
pASC1946	pASC18	vector	433					*Fsp*I
pASC1947	HI1502	H. influenzae predicted coding region HI1502	1575768	17	414	46.38	*12*	*Fsp*I
pASC1948	pASC18	vector	427					*Fsp*I
pASC2001	pASC18	vector	427					*Swa*I
pASC2002	pASC18	vector	427					*Swa*I
pASC2003	pASC18	vector	427					*Swa*I
pASC2004	pASC18	vector	427					*Swa*I
pASC2004	pASC18	vector	433					*Swa*I
pASC2005	pASC18	vector	427					*Swa*I
pASC2006	pASC18	vector	427					*Swa*I
pASC2007	pASC18	vector	433					*Swa*I
pASC2008	pASC18	vector	427					*Swa*I
pASC2009	pASC18	vector	433					*Swa*I
pASC2010	pASC18	vector	427					*Swa*I
pASC2011	pASC18	vector	427					*Swa*I
pASC2012	pASC18	vector	427					*Swa*I
pASC2025	pASC18	vector	1599					*Xmn*I
pASC2025	rrn	rRNA operon	rrn					*Xmn*I
pASC2026	pASC18	vector	563					*Xmn*I
pASC2026	rrn	rRNA operon	rrn					*Xmn*I
pASC2027	pASC18	vector	476					*Xmn*I
pASC2028	rrn	rRNA operon	rrn					*Xmn*I
pASC2029	rrn	rRNA operon	rrn					*Xmn*I
pASC2030	pASC18	vector	1242					*Xmn*I
pASC2031	rrn	rRNA operon	rrn					*Xmn*I
pASC2032	rrn	rRNA operon	rrn					*Xmn*I
pASC2033	rrn	rRNA operon	rrn					*Xmn*I
pASC2034	HI0752	phosphoribosylformylglycinamidine synthase (purL)	816831	1288	1320	42.1		*Xmn*I
pASC2036	HI0861	virulence-associated protein (vacB)	912162	188	782	40.62		*Xmn*I
pASC2037	rrn	rRNA operon	rrn					*Xmn*I

pASC2040	pASC18	vector	563					*Xmn*I
pASC2040	rrn	rRNA operon	rrn					*Xmn*I
pASC2042	pASC18	vector	1280					*Xmn*I
pASC2043	pASC18	vector	1451					*Xmn*I
pASC2044	rrn	rRNA operon	rrn					*Xmn*I
pASC2045	HI0752	phosphoribosylformylglycinamidine synthase (purL)	813934	322	1320	42.1		*Xmn*I
pASC2046	rrn	rRNA operon	rrn					*Xmn*I
pASC2047	rrn	rRNA operon	rrn					*Xmn*I
pASC2048	HI0739	DNA polymerase III, alpha subunit (dnaE)	798258	453	1159	41.79		*Xmn*I
pASC2049	HI0752	phosphoribosylformylglycinamidine synthase (purL)	814054	362	1320	42.1	*256*	*Xmn*I
pASC2050	HI0726	nitrate/nitrite response regulator protein (narP)	778161	146	208	36.86	*64*	*Xmn*I
pASC2051	HI0286	aminotransferase	320128	108	404	40.84		*Xmn*I
pASC2052	pASC18	vector	1551					*Xmn*I
pASC2053	HI1047	anaerobic dimethyl sulfoxide reductase, chain A (dmsA)	1111231	770	806	38.54	*219*	*Xmn*I
pASC2054	rrn	rRNA operon	rrn					*Xmn*I
pASC2055	rrn	rRNA operon	rrn					*Xmn*I
pASC2056	rrn	rRNA operon	rrn					*Xmn*I
pASC2057	pASC18	vector	563					*Xmn*I
pASC2057	pASC18	vector	1601					*Xmn*I
pASC2059	HI0080	conserved hypothetical protein	86803	76	143	31.93	*59*	*Xmn*I
pASC2060	HI1516	phage related protein frameshift mutation	1587368	129	419			*Xmn*I
pASC2061	HI1285	type I restriction enzyme (hsdR)	1364550	604	1055	43.13		*Xmn*I
pASC2062	pASC18	vector	1677					*Xmn*I
pASC2063	rrn	rRNA operon	rrn					*Xmn*I
pASC2064	pASC18	vector	1249					*Xmn*I
pASC2065	rrn	rRNA operon	rrn					*Xmn*I
pASC2066	rrn	rRNA operon	rrn					*Xmn*I
pASC2067	pASC18	vector	2141					*Xmn*I
pASC2067	rrn	rRNA operon	rrn					*Xmn*I
pASC2068	pASC18	vector	1823					*Xmn*I
pASC2069	pASC18	vector	1619					*Xmn*I
pASC2070	pASC18	vector	1242					*Xmn*I
pASC2071	pASC18	vector	939					*Xmn*I
pASC2071	pASC18	vector	1258					*Xmn*I
pASC2072	HI0751	thiol peroxidase (tpx)	812049	113	165	39.8		*Xmn*I

pASC2072	HI0752	phosphoribosylformylglycinamidine synthase (purL)	812735	232	1320	42.1		*Xmn*I	Intergenic
pASC2073	rrn	rRNA operon	rrn					*Xmn*I	
pASC2074	pASC18	vector	1838					*Xmn*I	
pASC2075	rrn	rRNA operon	rrn					*Xmn*I	
pASC2076	HI0752	phosphoribosylformylglycinamidine synthase (purL)	814153	395	1320	42.1	*157*	*Xmn*I	
pASC2077	pASC18	vector	440					*Xmn*I	Hinf junct @ 662461
pASC2078	pASC18	vector	1014					*Xmn*I	
pASC2079	rrn	rRNA operon	rrn					*Xmn*I	
pASC2080	pASC18	vector	1314					*Xmn*I	
pASC2080	HI1502	H. influenzae predicted coding region HI1502	1575768	17	414	46.38		*Xmn*I	
pASC2081	rrn	rRNA operon	rrn					*Xmn*I	
pASC2082	pASC18	vector	1866					*Xmn*I	
pASC2083	rrn	rRNA operon	rrn					*Xmn*I	
pASC2085	HI1502	H. influenzae predicted coding region HI1502	1575768	17	414	46.38		*Xmn*I	
pASC2086	pASC18	vector	2183					*Xmn*I	
pASC2087	HI0675	aminoacyl-histidine dipeptidase (pepD)	717893	208	484	40.36		*Xmn*I	
pASC2088	rrn	rRNA operon	rrn					*Xmn*I	
pASC2089	rrn	rRNA operon	rrn					*Xmn*I	
pASC2090	HI0231	ATP-dependent RNA helicase (deaD)	261885	104	613	42.03		*Xmn*I	
pASC2090	rrn	rRNA operon	rrn					*Xmn*I	
pASC2090	rrn	rRNA operon	rrn					*Xmn*I	
pASC2091	rrn	rRNA operon	rrn					*Xmn*I	
pASC2092	pASC18	vector	515					*Xmn*I	
pASC2093	HI1103	cysteine synthetase (cysK)	1166263	177	316	43.46	*266*	*Xmn*I	
pASC2094	pASC18	vector	1061					*Xmn*I	
pASC2095	pASC18	vector	625					*Xmn*I	
pASC2096	pASC18	vector	868					*Xmn*I	
pASC2097	HI0751	thiol peroxidase (tpx)	812211	59	165	39.8		*Xmn*I	
pASC2098	rrn	rRNA operon	rrn					*Xmn*I	
pASC2099	rrn	rRNA operon	rrn					*Xmn*I	
pASC2100	HI0339	primosomal protein N' (priA)	366589	513	730	41.28	**44**	*Xmn*I	
pASC2101	rrn	rRNA operon	rrn					*Xmn*I	
pASC2102	HI0921	leucyl-tRNA-synthetase (leuS)	977544	133	861	41.35		*Xmn*I	
pASC2103	HI1502	H. influenzae predicted coding region HI1502	1575768	17	414	46.38		*Xmn*I	
pASC2104	rrn	rRNA operon	rrn					*Xmn*I	
pASC2105	rrn	rRNA operon	rrn					*Xmn*I	

pASC2106	rrn	rRNA operon	rrn					*Xmn*I	
pASC2107	rrn	rRNA operon	rrn					*Xmn*I	
pASC2110	pASC18	vector	1260					*Xmn*I	
pASC2111	HI0307	pyrroline-5-carboxylate reductase (proC)	341260	207	271	41.45		*Xmn*I	
pASC2112	pASC18	vector	1135					*Xmn*I	
pASC2113	HI0264	heme-hemopexin utilization protein A (hxuA)	297338	368	905	35.99	**44**	*Xmn*I	
pASC2114	HI1145	chorismate mutase / prephenate dehydratase (pheA)	1214400	10	385	34.29		*Xmn*I	
pASC2115	pASC18	vector	713					*Xmn*I	
pASC2116	HI1135	phospho-N-acetylmuramoyl-pentapeptide-transferase E (mraY)	1203186	219	360	40.19	*257*	XmnI	
pASC2116	rrn	rRNA operon	rrn					*Xmn*I	
pASC2117	rrn	rRNA operon	rrn					*Xmn*I	
pASC2118	rrn	rRNA operon	rrn					*Xmn*I	
pASC2119	rrn	rRNA operon	rrn					*Xmn*I	
pASC2120	HI0862	conserved hypothetical protein	912939	229	235	39.29	*7*	*Xmn*I	
pASC2120	HI0862	conserved hypothetical protein	913542	28	235	39.29		*Xmn*I	
pASC2121	HI1139	UDP-N-acetylmuramate--alanine ligase (murC)	1208120	220	475	41.12		*Xmn*I	
pASC2122	pASC18	vector	1026					*Xmn*I	
pASC2122	rrn	rRNA operon	rrn					*Xmn*I	
pASC2123	HI0317	aspartyl-tRNA-synthetase (aspS)	348765	254	588	41.95		*Xmn*I	
pASC2124	HI0140	N-acetylglucosamine-6-phosphate deacetylase (nagA)	155414	346	381	38.58		*Xmn*I	
pASC2125	pASC18	vector	690					*Xmn*I	
pASC2126	HI0739	DNA polymerase III, alpha subunit (dnaE)	797628	663	1159	41.79	*239*	*Xmn*I	
pASC2127	pASC18	vector	555					*Xmn*I	
pASC2128	HI1285	type I restriction enzyme (hsdR)	1365313	349	1055	43.13		*Xmn*I	
pASC2129	HI0861	virulence-associated protein (vacB)	912162	188	782	40.62		*Xmn*I	
pASC2130	HI0655	shikimate 5-dehydrogenase (aroE)	699029	169	272	41.05	*26*	*Xmn*I	
pASC2131	pASC18	vector	894					*Xmn*I	
pASC2132	rrn	rRNA operon	rrn					*Xmn*I	
pASC2133	rrn	rRNA operon	rrn					*Xmn*I	
pASC2134	rrn	rRNA operon	rrn					*Xmn*I	
pASC2135	pASC18	vector	868					*Xmn*I	
pASC2136	HI1285	type I restriction enzyme (hsdR)	1363118	77	1055	43.13	*140*	*Xmn*I	Intergenic
pASC2137	HI1335	cell division protein (ftsH)	1411529	269	635	41.99	*284*	*Xmn*I	
pASC2138	HI1285	type I restriction enzyme (hsdR)	1364650	570	1055	43.13		*Xmn*I	

pASC2139	pASC18	vector	966					*Xmn*I
pASC2140	HI0286	aminotransferase	320273	156	404	40.84		*Xmn*I
pASC2141	rrn	rRNA operon	rrn					*Xmn*I
pASC2142	pASC18	vector	1278					*Xmn*I
pASC2142	pASC18	vector	2261					*Xmn*I
pASC2143	HI0751	thiol peroxidase (tpx)	812024	121	165	39.8		*Xmn*I
pASC2144	HI0514	DNA-directed RNA polymerase, beta' chain (rpoC)	534264	181	1415	41.37	*9*	*Xmn*I
pASC2145	pASC18	vector	433					*Xmn*I
pASC2146	HI0988	3-isopropylmalate dehydratase, alpha subunit (leuC)	1046826	195	469	43.64		*Xmn*I
pASC2147	pASC18	vector	524					*Xmn*I
pASC2148	HI1279	acylneuraminate cytidylyltransferase (neuA)	1356122	46	228	38.74	*265*	*Xmn*I
pASC2149	HI0078	cysteinyl-tRNA-synthetase (cysS)	85639	106	459	40.38		*Xmn*I
pASC2150	HI0624	sun protein (sun)	664703	178	451	40.95		*Xmn*I
pASC2151	HI0708	L-seryl-tRNA-selenium transferase (selA)	755057	368	461	39.26	*104*	*Xmn*I
pASC2152	HI1516	phage related protein frameshift mutation	1588039	353	419			*Xmn*I
pASC2153	HI0863	pyridoxamine phosphate oxidase (pdxH)	914432	223	229	41.19		*Xmn*I
pASC2154	pASC18	vector	1282					*Xmn*I
pASC2155	pASC18	vector	1271					*Xmn*I
pASC2156	HI0751	thiol peroxidase (tpx)	812124	88	165	39.8		*Xmn*I
pASC2157	HI0078	cysteinyl-tRNA-synthetase (cysS)	84973	328	459	40.38		*Xmn*I
pASC2158	HI1023	transketolase 1 (tktA)	1087945	438	665	43.06		*Xmn*I
pASC2159	HI0859	ATP-dependent Clp protease, ATPase subunit (clpB)	908356	460	856	41.16		*Xmn*I
pASC2160	pASC18	vector	1348					*Xmn*I
pASC2161	HI0752	phosphoribosylformylglycinamidine synthase (purL)	815499	844	1320	42.1	*189*	*Xmn*I
pASC2162	pASC18	vector	849					*Xmn*I
pASC2163	HI1675	molybdenum cofactor biosynthesis protein C (moaC)	1743963	56	160	42.08	*99*	*Xmn*I
pASC2163	rrn	rRNA operon	rrn					*Xmn*I
pASC2164	HI0309	integrase/recombinase (xerD)	343114	155	297	42.09		*Xmn*I
pASC2165	HI0078	cysteinyl-tRNA-synthetase (cysS)	85356	201	459	40.38		*Xmn*I
pASC2166	rrn	rRNA operon	rrn					*Xmn*I
pASC2167	pASC18	vector	1301					*Xmn*I
pASC2168	pASC18	vector	849					*Xmn*I
pASC2169	HI0469	histidinol dehydrogenase (hisD)	492023	129	427	42.31	*107*	*Xmn*I
pASC2169	HI1287	type I modification enzyme (hsdM)	1369124	11	443	48.76	*47*	*Xmn*I
pASC2170	rrn	rRNA operon	rrn					*Xmn*I
pASC2171	HI0750	diaminopimelate epimerase (dapF)	811542	231	274	39.78	*143*	*Xmn*I

pASC2171	HI0750	diaminopimelate epimerase (dapF)	811657	269	274	39.78	*258*	*Xmn*I
pASC2172	pASC18	vector	749					*Xmn*I
pASC2173	pASC18	vector	527					*Xmn*I
pASC2174	pASC18	vector	483					*Xmn*I
pASC2176	rrn	rRNA operon	rrn					*Xmn*I
pASC2177	pASC18	vector	1028					*Xmn*I
pASC2178	pASC18	vector	1028					*Xmn*I
pASC2179	pASC18	vector	1249					*Xmn*I
pASC2180	pASC18	vector	433					*Xmn*I
pASC2181	pASC18	vector	433					*Xmn*I
pASC2182	pASC18	vector	1540					*Xmn*I
pASC2183	HI1287	type I modification enzyme (hsdM)	1368964	64	443	48.76	*113*	*Xmn*I
pASC2184	HI1520	conserved hypothetical protein	1589965	241	355	48.26		*Xmn*I
pASC2185	rrn	rRNA operon	rrn					*Xmn*I
pASC2186	pASC18	vector	542					*Xmn*I
pASC2187	HI1287	type I modification enzyme (hsdM)	1368283	291	443	48.76		*Xmn*I
pASC2188	pASC18	vector	1529					*Xmn*I
pASC2189	rrn	rRNA operon	rrn					*Xmn*I
pASC2190	HI0752	phosphoribosylformylglycinamidine synthase (purL)	813502	178	1320	42.1		*Xmn*I
pASC2191	pASC18	vector	1271					*Xmn*I
pASC2192	pASC18	vector	2008					*Xmn*I
pASC2194	rrn	rRNA operon	rrn					*Xmn*I
pASC2195	HI1103	cysteine synthetase (cysK)	1166247	171	316	43.46	*282*	*Xmn*I
pASC2196	HI1285	type I restriction enzyme (hsdR)	1364442	640	1055	43.13		*Xmn*I
pASC2196	HI1287	type I modification enzyme (hsdM)	1368269	296	443	48.76		*Xmn*I
pASC2198	pASC18	vector	894					*Xmn*I
pASC2199	pASC18	vector	1258					*Xmn*I
pASC2200	rrn	rRNA operon	rrn					*Xmn*I
pASC2201	HI0316	datP pyrophosphohydrolase (ntpA)	347441	100	158	36.71	**55**	*Xmn*I
pASC2202	HI0625	TRK system potassium uptake protein (trkA)	665608	24	458	38.86	*274*	*Xmn*I
pASC2204	HI1287	type I modification enzyme (hsdM)	1368330	275	443	48.76		*Xmn*I
pASC2205	HI1465	cell division ftsH-related protein	1549880	82	381	43.13		*Xmn*I
pASC2207	rrn	rRNA operon	rrn					*Xmn*I
pASC2208	HI1572	rcb point mutation	1638651	358	365		*148*	*Xmn*I
pASC2210	HI0752	phosphoribosylformylglycinamidine synthase (purL)	813557	196	1320	42.1		*Xmn*I
pASC2211	HI0861	virulence-associated protein (vacB)	912162	188	782	40.62		*Xmn*I

pASC2212	pASC18	vector	810					*Xmn*I
pASC2214	pASC18	vector	1784					*Xmn*I
pASC2215	HI0752	phosphoribosylformylglycinamidine synthase (purL)	816226	1086	1320	42.1		*Xmn*I
pASC2216	pASC18	vector	1528					*Xmn*I
pASC2217	HI1042	metB frameshift mutation	1108049	91	616			*Xmn*I
pASC2218	pASC18	vector	1277					*Xmn*I
pASC2218	HI0140	N-acetylglucosamine-6-phosphate deacetylase (nagA)	155414	346	381	38.58		*Xmn*I
pASC2219	HI1444	5,10 methylenetetrahydrofolate reductase (metF)	1532091	81	292	36.76	*134*	*Xmn*I
pASC2220	pASC18	vector	1277					*Xmn*I
pASC2221	pASC18	vector	2188					*Xmn*I
pASC2222	rrn	rRNA operon	rrn					*Xmn*I
pASC2223	HI0921	leucyl-tRNA-synthetase (leuS)	978429	428	861	41.35		*Xmn*I
pASC2224	HI1311	phenylalanyl-tRNA-synthetase, alpha subunit (pheS)	1389132	277	329	39.11		*Xmn*I
pASC2225	HI0752	phosphoribosylformylglycinamidine synthase (purL)	815272	768	1320	42.1	*38*	*Xmn*I
pASC2226	HI1374	cell division protein (mukB)	1465935	32	1510	39.96		*Xmn*I
pASC2227	HI0078	cysteinyl-tRNA-synthetase (cysS)	85333	208	459	40.38		*Xmn*I
pASC2228	pASC18	vector	1208					*Xmn*I
pASC2229	rrn	rRNA operon	rrn					*Xmn*I
pASC2230	rrn	rRNA operon	rrn					*Xmn*I
pASC2231	pASC18	vector	524					*Xmn*I
pASC2232	HI1287	type I modification enzyme (hsdM)	1368672	161	443	48.76		*Xmn*I
pASC2233	pASC18	vector	850					*Xmn*I
pASC2233	pASC18	vector	2033					*Xmn*I
pASC2234	HI0752	phosphoribosylformylglycinamidine synthase (purL)	814026	353	1320	42.1	*284*	*Xmn*I
pASC2235	rrn	rRNA operon	rrn					*Xmn*I
pASC2236	pASC18	vector	902					*Xmn*I
pASC2237	pASC18	vector	542					*Xmn*I
pASC2238	pASC18	vector	1242					*Xmn*I
pASC2239	rrn	rRNA operon	rrn					*Xmn*I
pASC2240	rrn	rRNA operon	rrn					*Xmn*I
pASC2241	pASC18	vector	1222					*Xmn*I
pASC2241	HI0579	elongation factor G (fusA)	598347	500	700	40.67	*132*	*Xmn*I
pASC2242	rrn	rRNA operon	rrn					*Xmn*I
pASC2243	HI0861	virulence-associated protein (vacB)	912162	188	782	40.62		*Xmn*I
pASC2244	pASC18	vector	1215					*Xmn*I
pASC2245	pASC18	vector	1242					*Xmn*I

pASC2246	HI0752	phosphoribosylformylglycinamidine synthase (purL)	815478	837	1320	42.1	*168*	*Xmn*I	
pASC2247	pASC18	vector	754					*Xmn*I	
pASC2248	HI0752	phosphoribosylformylglycinamidine synthase (purL)	814129	387	1320	42.1	*181*	*Xmn*I	
pASC2249	pASC18	vector	736					*Xmn*I	
pASC2249	HI0229	polynucleotide phosphorylase (pnp)	258846	183	709	42.5		*Xmn*I	
pASC2250	pASC18	vector	562					*Xmn*I	
pASC2251	HI0989	3-isopropylmalate dehydratase small subunit (leuD)	1047966	97	200	40		*Xmn*I	
pASC2252	pASC18	vector	758					*Xmn*I	
pASC2252	rrn	rRNA operon	rrn					*Xmn*I	
pASC2253	pASC18	vector	534					*Xmn*I	
pASC2254	pASC18	vector	1443					*Xmn*I	
pASC2255	pASC18	vector	555					*Xmn*I	
pASC2256	pASC18	vector	524					*Xmn*I	
pASC2257	pASC18	vector	913					*Xmn*I	
pASC2257	pASC18	vector	1673					*Xmn*I	
pASC2258	rrn	rRNA operon	rrn					*Xmn*I	
pASC2258	rrn	rRNA operon	rrn					*Xmn*I	
pASC2259	rrn	rRNA operon	rrn					*Xmn*I	
pASC2260	pASC18	vector	939					*Xmn*I	
pASC2261	pASC18	vector	710					*Xmn*I	
pASC2262	pASC18	vector	749					*Xmn*I	
pASC2263	rrn	rRNA operon	rrn					*Xmn*I	
pASC2264	rrn	rRNA operon	rrn					*Xmn*I	
pASC2265	HI0090	conserved hypothetical protein	96962	234	237	38.54		*Swa*I	Intergenic
pASC2266	HI0089	aspartokinase I / homoserine dehydrogenase I (thrA)	96442	143	815	40.57		*Swa*I	
pASC2267	HI0088	homoserine kinase (thrB)	93499	304	314	38.75	*39*	*Swa*I	
pASC2268	HI0913	ribosomal protein S2 (rpS2)	968083	65	251	42.23		*Swa*I	
pASC2269	HI1193	branched-chain-amino-acid transaminase (ilvE)	1258821	271	343	43.54		*Swa*I	
pASC2270	HI0412	conserved hypothetical protein	432765	31	322	38.61	*94*	*Swa*I	Intergenic
pASC2271	rrn	rRNA operon	rrn					*Swa*I	
pASC2272	HI0089	aspartokinase I / homoserine dehydrogenase I (thrA)	96466	135	815	40.57		*Swa*I	
pASC2273	HI0089	aspartokinase I / homoserine dehydrogenase I (thrA)	95987	295	815	40.57		*Swa*I	
pASC2274	pASC18	vector	845					*Swa*I	
pASC2274	pASC18	vector	1941					*Swa*I	
pASC2275	HI0525	phosphoglycerate kinase (pgk)	549532	1	386	40.24	*279*	*Swa*I	
pASC2276	pASC18	vector	443					*Swa*I	

pASC2277	HI1335	cell division protein (ftsH)	1412056	444	635	41.99		*Swa*I	
pASC2278	HI1094	cytochrome C-type biogenesis protein (ccmF)	1157190	193	648	41.51		*Swa*I	
pASC2279	HI0471	imidazoleglycerol-phosphate dehydratase / histidinol-phosphatase (hisB)	494879	225	362	40.79	*288*	*Swa*I	
pASC2280	pASC18	vector	2267					*Swa*I	
pASC2281	HI0089	aspartokinase I / homoserine dehydrogenase I (thrA)	96173	233	815	40.57		*Swa*I	
pASC2282	HI0616	ATP-dependent helicase (hepA)	649361	346	923	40.95		*Swa*I	
pASC2283	HI0471	imidazoleglycerol-phosphate dehydratase / histidinol-phosphatase (hisB)	494904	234	362	40.79	*263*	*Swa*I	
pASC2284	pASC18	vector	1270					*Swa*I	
pASC2285	HI1339	H. influenzae predicted coding region HI1339	1415899	67	129	37.47	*201*	*Swa*I	
pASC2286	HI1339	H. influenzae predicted coding region HI1339	1415899	67	129	37.47	*201*	*Swa*I	
pASC2287	HI1339	H. influenzae predicted coding region HI1339	1415899	67	129	37.47	*201*	*Swa*I	
pASC2288	HI1339	H. influenzae predicted coding region HI1339	1415899	67	129	37.47	*201*	*Swa*I	
pASC2289	HI0576	conserved hypothetical protein	595553	73	126	37.3		*Swa*I	
pASC2290	HI1430	short chain dehydrogenase/reductase	1518402	34	252	41.67		*Swa*I	Intergenic
pASC2291	pASC18	vector	2300					*Swa*I	
pASC2292	pASC18	vector	1240					*Swa*I	
pASC2292	HI0354	ABC transporter, ATP-binding protein	381174	177	240	46.94		*Swa*I	
pASC2292	HI0914	elongation factor Ts (tsf)	969326	183	283	39.81		*Swa*I	
pASC2293	HI0927	glycyl-tRNA-synthetase, alpha chain (glyQ)	984844	259	302	42.72		*Swa*I	
pASC2293	HI1339	H. influenzae predicted coding region HI1339	1415820	40	129	37.47	*280*	*Swa*I	
pASC2294	pASC18	vector	1278					*Swa*I	
pASC2295	pASC18	vector	91					*Swa*I	
pASC2296	HI0913	ribosomal protein S2 (rpS2)	968176	96	251	42.23		*Swa*I	
pASC2297	HI0006	formate dehydrogenase, alpha subunit (fdxG)	6552	298	1029			*Swa*I	
pASC2299	HI0089	aspartokinase I / homoserine dehydrogenase I (thrA)	96760	37	815	40.57		*Swa*I	
pASC2300	pASC18	vector	1232					*Swa*I	
pASC2301	HI0089	aspartokinase I / homoserine dehydrogenase I (thrA)	96173	233	815	40.57		*Swa*I	
pASC2302	HI0354	ABC transporter, ATP-binding protein	380996	117	240	46.94		*Swa*I	
pASC2303	HI1339	H. influenzae predicted coding region HI1339	1415820	40	129	37.47	*280*	*Swa*I	
pASC2304	HI0070	DNA repair protein (recN)	77169	196	558	37.75		*Swa*I	
pASC2305	HI0090	conserved hypothetical protein	97027	169	237	38.54		*Swa*I	Intergenic
pASC2306	HI1337	mrsA protein (mrsA)	1414070	154	445	41.42		*Swa*I	
pASC2306	HI1337	mrsA protein (mrsA)	1414418	270	445	41.42		*Swa*I	
pASC2306	HI1337	mrsA protein (mrsA)	1414425	273	445	41.42		*Swa*I	

pASC2307	pASC18	vector	1242					*Swa*I	
pASC2308	HI1730	conserved hypothetical protein	1803343	224	309	41.1	*260*	*Swa*I	
pASC2309	HI1051	ABC transporter, ATP-binding protein	1115982	505	614	39.74		*Swa*I	
pASC2310	HI0089	aspartokinase I / homoserine dehydrogenase I (thrA)	96442	143	815	40.57		*Swa*I	
pASC2311	HI1511	sheath protein gpL (muL)	1581592	60	487	47.98		*Swa*I	
pASC2312	rrn	rRNA operon	rrn					*Swa*I	
pASC2313	HI1574	replicative DNA helicase (dnaB)	1643040	446	504	40.01		*Swa*I	
pASC2314	rrn	rRNA operon	rrn					*Swa*I	
pASC2315	pASC18	vector	1270					*Swa*I	
pASC2316	HI0089	aspartokinase I / homoserine dehydrogenase I (thrA)	95950	307	815	40.57		*Swa*I	
pASC2317	HI0914	elongation factor Ts (tsf)	969326	183	283	39.81		*Swa*I	
pASC2318	HI0469	histidinol dehydrogenase (hisD)	491862	75	427	42.31	*291*	*Swa*I	
pASC2319	rrn	rRNA operon	rrn					*Swa*I	
pASC2320	pASC18	vector	1232					*Swa*I	
pASC2321	pASC18	vector	1242					*Swa*I	
pASC2322	HI0528	tyrosine-specific transport protein (tyrP)	551548	202	406	41.54		*Swa*I	
pASC2323	HI1575	alanine racemase, biosynthetic (alr)	1643973	249	360	40.65	**125**	*Swa*I	
pASC2324	HI0090	conserved hypothetical protein	97027	169	237	38.54		*Swa*I	Intergenic
pASC2325	HI0090	conserved hypothetical protein	96975	221	237	38.54		*Swa*I	Intergenic
pASC2326	HI0070	DNA repair protein (recN)	77038	240	558	37.75		*Swa*I	
pASC2327	HI0088	homoserine kinase (thrB)	93499	304	314	38.75	*39*	*Swa*I	
pASC2329	HI1574	replicative DNA helicase (dnaB)	1642127	141	504	40.01		*Swa*I	
pASC2330	HI0089	aspartokinase I / homoserine dehydrogenase I (thrA)	96760	37	815	40.57		*Swa*I	
pASC2331	HI0090	conserved hypothetical protein	97027	169	237	38.54		*Swa*I	Intergenic
pASC2332	HI1339	H. influenzae predicted coding region HI1339	1415899	67	129	37.47	*201*	*Swa*I	
pASC2333	rrn	rRNA operon	rrn					*Swa*I	
pASC2334	rrn	rRNA operon	rrn					*Swa*I	
pASC2335	HI0089	aspartokinase I / homoserine dehydrogenase I (thrA)	95802	356	815	40.57		*Swa*I	
pASC2336	pASC18	vector	1249					*Swa*I	
pASC2337	pASC18	vector	2176					*Swa*I	
pASC2338	HI0090	conserved hypothetical protein	97027	169	237	38.54		*Swa*I	Intergenic
pASC2339	HI0090	conserved hypothetical protein	97027	169	237	38.54		*Swa*I	Intergenic
pASC2340	HI0927	glycyl-tRNA-synthetase, alpha chain (glyQ)	985055	189	302	42.72		*Swa*I	
pASC2341	HI0528	tyrosine-specific transport protein (tyrP)	551548	202	406	41.54		*Swa*I	
pASC2342	HI1295	nifS protein, putative	1374638	338	437	41.11		*Swa*I	
pASC2343	pASC18	vector	1247					*Swa*I	

pASC2344	HI1617	aspartate aminotransferase (aspC)	1684735	21	396	39.65	*81*	*Swa*I	Intergenic
pASC2345	HI0089	aspartokinase I / homoserine dehydrogenase I (thrA)	95950	307	815	40.57		*Swa*I	
pASC2346	HI1193	branched-chain-amino-acid transaminase (ilvE)	1259402	77	343	43.54	*299*	*Swa*I	
pASC2347	HI0525	phosphoglycerate kinase (pgk)	549439	32	386	40.24	*186*	*Swa*I	
pASC2348	HI1574	replicative DNA helicase (dnaB)	1643089	462	504	40.01		*Swa*I	
pASC2349	HI0089	aspartokinase I / homoserine dehydrogenase I (thrA)	96172	233	815	40.57		*Swa*I	
pASC2350	pASC18	vector	263					*Swa*I	
pASC2350	HI0089	aspartokinase I / homoserine dehydrogenase I (thrA)	96442	143	815	40.57		*Swa*I	
pASC2351	HI0090	conserved hypothetical protein	96962	234	237	38.54		*Swa*I	Intergenic
pASC2352	HI1732	adhesin frameshift	1806191	889	1323			*Swa*I	
pASC2353	HI0525	phosphoglycerate kinase (pgk)	549532	1	386	40.24	*279*	*Swa*I	
pASC2354	HI1574	replicative DNA helicase (dnaB)	1643089	462	504	40.01		*Swa*I	
pASC2355	HI0090	conserved hypothetical protein	96962	234	237	38.54		*Swa*I	Intergenic
pASC2356	HI0090	conserved hypothetical protein	97053	143	237	38.54		*Swa*I	Intergenic
pASC2357	HI0469	histidinol dehydrogenase (hisD)	491902	89	427	42.31		*Swa*I	
pASC2358	pASC18	vector	2177					*Swa*I	
pASC2359	pASC18	vector	1242					*Swa*I	
pASC2360	HI1339	H. influenzae predicted coding region HI1339	1415857	53	129	37.47	*243*	*Swa*I	
pASC2362	pASC18	vector	1232					*Swa*I	
pASC2363	HI0216	type I restriction/modification specificity protein (hsdS)	233876	133	385	34.89		*Swa*I	
pASC2364	HI0914	elongation factor Ts (tsf)	969457	227	283	39.81		*Swa*I	
pASC2365	HI0913	ribosomal protein S2 (rpS2)	968083	65	251	42.23		*Swa*I	
pASC2366	HI1467	ABC transporter, ATP-binding protein	1553931	468	589	34.92	*181*	*Swa*I	
pASC2366	rrn	rRNA operon	rrn					*Swa*I	
pASC2367	HI0089	aspartokinase I / homoserine dehydrogenase I (thrA)	96173	233	815	40.57		*Swa*I	
pASC2368	HI0090	conserved hypothetical protein	97027	169	237	38.54		*Swa*I	Intergenic
pASC2369	HI0576.1	conserved hypothetical protein	595928	72	119	33.33	**44**	*Swa*I	
pASC2370	HI0088	homoserine kinase (thrB)	93479	311	314	38.75	*19*	*Swa*I	
pASC2371	pASC18	vector	2177					*Swa*I	
pASC2372	HI0088	homoserine kinase (thrB)	93492	306	314	38.75	*32*	*Swa*I	
pASC2373	HI0089	aspartokinase I / homoserine dehydrogenase I (thrA)	95885	329	815	40.57		*Swa*I	
pASC2374	pASC18	vector	147					*Swa*I	
pASC2374	HI0089	aspartokinase I / homoserine dehydrogenase I (thrA)	96299	191	815	40.57		*Swa*I	
pASC2375	HI0469	histidinol dehydrogenase (hisD)	491926	97	427	42.31		*Swa*I	
pASC2376	HI0089	aspartokinase I / homoserine dehydrogenase I (thrA)	96442	143	815	40.57		*Swa*I	

pASC2377	HI0088	homoserine kinase (thrB)	94331	27	314	38.75		*Swa*I
pASC2378	HI0088	homoserine kinase (thrB)	93809	201	314	38.75		*Swa*I
pASC2378	HI1335	cell division protein (ftsH)	1411576	284	635	41.99		*Swa*I
pASC2379	pASC18	vector	166					*Swa*I
pASC2380	pASC18	vector	1210					*Swa*I
pASC2381	HI1576	glucose-6-phosphate isomerase (pgi)	1644704	113	563	36.71		*Swa*I
pASC2382	pASC18	vector	2300					*Swa*I
pASC2382	pASC18	vector	2307					*Swa*I
pASC2383	HI1735	peptide chain release factor 3 (prfC)	1812590	115	527	41.68		*Swa*I
pASC2384	HI0088	homoserine kinase (thrB)	93479	311	314	38.75	*19*	*Swa*I

Gray shading indicates multiple insertions mapped in the same sequencing reaction

† Distance of Tn5 insertion from insert/vector junction (**bold** = experimental; *italics* = theoretical)

Appendix E – Mapped Tn5 insertions (pASC18, post-mutagenesis minimalization)

Allele/Plasmid	CDS/Vector	Protein/Gene	Mutation coordinate	Disrupted codon	Total codons	Percent GC	Distance from junct (<300nt) †	Library	Comments
pASC2385	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007706	153	511	42.14		*Psi*I	
pASC2386	pASC18	vector	365					*Psi*I	
pASC2387	pASC18	vector	579					*Psi*I	
pASC2388	rrn	rRNA operon	rrn					*Psi*I	
pASC2389	HI0216	type I restriction/modification specificity protein (hsdS)	233899	141	385	34.89	*267*	*Psi*I	
pASC2389	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1006968	399	511	42.14		*Psi*I	
pASC2390	HI1120	oligopeptide ABC transporter, ATP-binding protein (oppF)	1186416	323	332	41.37	**27**	*Psi*I	
pASC2391	HI0588	N-carbamyl-L-amino acid amidohydrolase	610340	366	411	34.23	*142*	*Psi*I	
pASC2391	HI1374	cell division protein (mukB)	1467176	446	1510	39.96		*Psi*I	
pASC2392	pASC18	vector	556					*Psi*I	
pASC2393	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007834	110	511	42.14		*Psi*I	
pASC2394	HI0567	DNA gyrase, subunit B (gyrB)	587582	138	806	41.11	*77*	*Psi*I	
pASC2394	HI0567	DNA gyrase, subunit B (gyrB)	587589	135	806	41.11	*70*	*Psi*I	
pASC2395	rrn	rRNA operon	rrn					*Psi*I	
pASC2396	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007731	144	511	42.14		*Psi*I	
pASC2397	HI0216	type I restriction/modification specificity protein (hsdS)	234006	176	385	34.89	*160*	*Psi*I	
pASC2398	HI0897	multidrug resistance protein B (emrB)	953184	327	510	39.67	**17**	*Psi*I	
pASC2399	HI1615	phosphoribosylaminoimidazole carboxylase, catalytic subunit (purE)	1683389	141	164	45.12		*Psi*I	
pASC2400	HI1516	phage related protein frameshift mutation	1587643	221	419			*Psi*I	
pASC2401	HI1262	sanA protein (sanA)	1340047	146	244	39.48	*293*	*Psi*I	
pASC2402	HI0216	type I restriction/modification specificity protein (hsdS)	233983	168	385	34.89	*183*	*Psi*I	
pASC2403	HI0229	polynucleotide phosphorylase (pnp)	259598	434	709	42.5	*166*	*Psi*I	
pASC2404	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007601	188	511	42.14		*Psi*I	
pASC2405	HI0729	prolyl-tRNA-synthetase (proS)	783614	349	572	43.12		*Psi*I	
pASC2406	HI1335	cell division protein (ftsH)	1412066	448	635	41.99		*Psi*I	
pASC2407	HI1520	conserved hypothetical protein	1589263	7	355	48.26		*Psi*I	

pASC2408	HI1518	H. influenzae predicted coding region HI1518	1588272	5	182	50.18		*Psi*I	Intergenic
pASC2409	pASC18	vector	433					*Psi*I	
pASC2410	pASC18	vector	2228					*Psi*I	
pASC2410	HI0947	virulence associated protein C (vapC)	1008583	0	132	37.12		*Psi*I	
pASC2412	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007767	132	511	42.14		*Psi*I	
pASC2413	pASC18	vector	584					*Psi*I	
pASC2414	HI1516	phage related protein frameshift mutation	1587997	339	419			*Psi*I	
pASC2415	HI0947	virulence associated protein C (vapC)	1008583	0	132	37.12		*Psi*I	
pASC2416	HI0091	conserved hypothetical protein	98055	343	378	44.62		*Psi*I	
pASC2416	HI1522	H. influenzae predicted coding region HI1522	1591079	59	623	41.63	*102*	*Psi*I	
pASC2418	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007423	247	511	42.14		*Psi*I	
pASC2419	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007348	272	511	42.14		*Psi*I	
pASC2420	rrn	rRNA operon	rrn					*Psi*I	
pASC2421	HI1515	64 kDa virion protein (muN)	1585883	90	455	46.15		*Psi*I	
pASC2422	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1008085	26	511	42.14		*Psi*I	
pASC2422	HI0949	aminotransferase	1009284	347	454	43.1		*Psi*I	
pASC2423	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007990	58	511	42.14		*Psi*I	
pASC2424	HI0062	dnaK suppressor protein (dksA)	66021	66	145	40.69		*Psi*I	
pASC2424	rrn	rRNA operon	rrn					*Psi*I	
pASC2425	HI0559.1	H. influenzae predicted coding region HI0559.1	579549	43	115	39.42	**72**	*Psi*I	
pASC2426	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007912	84	511	42.14		*Psi*I	
pASC2433	rrn	rRNA operon	rrn					*Psi*I	
pASC2434	HI1514	H. influenzae predicted coding region HI1514	1583941	74	631	48.71	*217*	*Psi*I	
pASC2435	HI1501	H. influenzae predicted coding region HI1501	1574420	119	520	48.33		*Psi*I	
pASC2435	HI1518	H. influenzae predicted coding region HI1518	1588387	36	182	50.18		*Psi*I	
pASC2436	rrn	rRNA operon	rrn					*Psi*I	
pASC2437	HI0214	oligopeptidase A (prlC)	229979	127	681	40.82	*59*	*Psi*I	
pASC2437	HI0216	type I restriction/modification specificity protein (hsdS)	233983	168	385	34.89	*183*	*Psi*I	
pASC2437	HI0947	virulence associated protein C (vapC)	1008583	0	132	37.12		*Psi*I	
pASC2438	HI0216	type I restriction/modification specificity protein (hsdS)	233849	124	385	34.89		*Psi*I	
pASC2439	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007974	63	511	42.14		*Psi*I	
pASC2440	pASC18	vector	1280					*Psi*I	
pASC2441	HI1515	64 kDa virion protein (muN)	1586525	304	455	46.15		*Psi*I	
pASC2442	HI1516	phage related protein frameshift mutation	1587650	223	419			*Psi*I	
pASC2442	HI1520	conserved hypothetical protein	1589892	217	355	48.26		*Psi*I	
pASC2443	HI1514	H. influenzae predicted coding region HI1514	1584989	423	631	48.71	*156*	*Psi*I	

pASC2443	rrn	rRNA operon	rrn					*Psi*I	
pASC2445	pASC18	vector	629					*Psi*I	
pASC2445	HI1515	64 kDa virion protein (muN)	1586040	142	455	46.15		*Psi*I	
pASC2446	pASC18	vector	1239					*Psi*I	
pASC2446	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007162	334	511	42.14		*Psi*I	
pASC2447	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007781	128	511	42.14		*Psi*I	
pASC2448	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007957	69	511	42.14		*Psi*I	
pASC2449	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007988	59	511	42.14		*Psi*I	
pASC2450	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007423	247	511	42.14		*Psi*I	
pASC2451	rrn	rRNA operon	rrn					*Psi*I	
pASC2452	pASC18	vector	170					*Psi*I	
pASC2453	pASC18	vector	418					*Psi*I	Hinf junct @ 1584832
pASC2454	HI0646	aspartate-semialdehyde dehydrogenase (asd)	690781	187	371	40.52		*Psi*I	
pASC2455	HI0994	transferrin-binding protein 1 precursor (tbp1)	1058829	458	912	35.86	**43**	*Psi*I	
pASC2456	HI1520	conserved hypothetical protein	1589800	186	355	48.26		*Psi*I	
pASC2456	HI1521	conserved hypothetical protein	1590689	98	168	41.07		*Psi*I	
pASC2457	pASC18	vector	1249					*Psi*I	
pASC2458	HI0020	transport protein, putative	22362	108	479	40.85	*67*	*Psi*I	
pASC2461	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007975	63	511	42.14		*Psi*I	
pASC2462	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007433	244	511	42.14		*Psi*I	
pASC2463	HI1282	conserved hypothetical protein	1357614	1072	141	39.72		*Psi*I	Intergenic
pASC2464	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007513	217	511	42.14		*Psi*I	
pASC2466	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1006885	426	511	42.14		*Psi*I	
pASC2467	HI1255	H. influenzae predicted coding region HI1255	1331037	186	74	45.5	*127*	*Psi*I	Intergenic
pASC2468	HI0213	oligopeptide transporter, periplasmic-binding protein, putative	227512	65	514	35.99		*Psi*I	Intergenic
pASC2469	HI1519	H. influenzae predicted coding region HI1519	1588931	32	135	49.88		*Psi*I	
pASC2469	HI1519	H. influenzae predicted coding region HI1519	1588938	34	135	49.88		*Psi*I	
pASC2470	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007649	172	511	42.14		*Psi*I	
pASC2471	pASC18	vector	2228					*Psi*I	
pASC2471	HI0947	virulence associated protein C (vapC)	1008583	0	132	37.12		*Psi*I	
pASC2472	pASC18	vector	433					*Psi*I	
pASC2473	HI1042	metB frameshift mutation	1107590	244	616			*Psi*I	
pASC2473	HI1518	H. influenzae predicted coding region HI1518	1588306	9	182	50.18		*Psi*I	
pASC2474	pASC18	vector	2274					*Psi*I	
pASC2475	HI1610	tyrosyl tRNA-synthetase (tyrS)	1676830	105	401	39.9	*10*	*Psi*I	

pASC2476	rrn	rRNA operon	rrn					*Psi*I	
pASC2477	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007423	247	511	42.14		*Psi*I	
pASC2478	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007487	226	511	42.14		*Psi*I	
pASC2479	rrn	rRNA operon	rrn					*Psi*I	
pASC2480	HI1521	conserved hypothetical protein	1590355	39	168	41.07		*Psi*I	Intergenic
pASC2493	pASC18	vector	587					*Psi*I	
pASC2494	pASC18	vector	1242					*Psi*I	
pASC2494	HI0479	ATP synthase F1, subunit beta (atpD)	500737	303	457	40.12	*184*	*Psi*I	
pASC2495	pASC18	vector	1273					*Psi*I	
pASC2495	rrn	rRNA operon	rrn					*Psi*I	
pASC2496	HI0479	ATP synthase F1, subunit beta (atpD)	501198	149	457	40.12		*Psi*I	
pASC2496	rrn	rRNA operon	rrn					*Psi*I	
pASC2497	HI0477	tyrosine-specific transport protein (tyrP)	498601	58	400	39.33	*249*	*Psi*I	
pASC2498	HI0543	heat shock protein (groEL)	567297	450	548	41.91	*86*	*Psi*I	
pASC2498	HI1516	phage related protein frameshift mutation	1588028	349	419			*Psi*I	
pASC2499	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007711	151	511	42.14		*Psi*I	
pASC2500	rrn	rRNA operon	rrn					*Psi*I	
pASC2501	HI0091	conserved hypothetical protein	98939	48	378	44.62		*Psi*I	
pASC2501	HI0355	ABC transporter, permease protein	381546	61	245	46.39		*Psi*I	
pASC2502	HI0729	prolyl-tRNA-synthetase (proS)	784049	494	572	43.12		*Psi*I	
pASC2503	HI0479	ATP synthase F1, subunit beta (atpD)	500694	317	457	40.12	*141*	*Psi*I	
pASC2504	HI1520	conserved hypothetical protein	1589385	48	355	48.26		*Psi*I	
pASC2505	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007556	203	511	42.14		*Psi*I	
pASC2506	rrn	rRNA operon	rrn					*Psi*I	
pASC2507	rrn	rRNA operon	rrn					*Psi*I	
pASC2508	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007622	181	511	42.14		*Psi*I	
pASC2509	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007394	257	511	42.14		*Psi*I	
pASC2510	pASC18	vector	564					*Psi*I	
pASC2510	HI0216	type I restriction/modification specificity protein (hsdS)	233900	141	385	34.89	*266*	*Psi*I	
pASC2511	HI1516	phage related protein frameshift mutation	1588087	369	419			*Psi*I	
pASC2511	HI1518	H. influenzae predicted coding region HI1518	1588662	128	182	50.18		*Psi*I	
pASC2513	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1008149	5	511	42.14		*Psi*I	
pASC2514	rrn	rRNA operon	rrn					*Psi*I	
pASC2515	HI0948	conserved hypothetical protein	1008776	13	77	39.83		*Psi*I	
pASC2515	HI1145	chorismate mutase / prephenate dehydratase (pheA)	1214480	37	385	34.29	*125*	*Psi*I	
pASC2516	HI1368	zinc protease, putative	1456671	882	926	32.79	**129**	*Psi*I	

pASC2517	rrn	rRNA operon	rrn					*Psi*I	
pASC2518	HI1610	tyrosyl tRNA-synthetase (tyrS)	1676603	180	401	39.9	*217*	*Psi*I	
pASC2520	HI0898	multidrug resistance protein A (emrA)	954637	237	390	38.72	*279*	*Psi*I	
pASC2520	rrn	rRNA operon	rrn					*Psi*I	
pASC2521	pASC18	vector	2300					*Psi*I	
pASC2522	HI0216	type I restriction/modification specificity protein (hsdS)	233849	124	385	34.89		*Psi*I	
pASC2522	HI0928	catalase (hktE)	987002	371	508	42.72		*Psi*I	
pASC2522	rrn	rRNA operon	rrn					*Psi*I	
pASC2523	HI1520	conserved hypothetical protein	1590110	289	355	48.26		*Psi*I	
pASC2524	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007610	185	511	42.14		*Psi*I	
pASC2525	HI0213	oligopeptide transporter, periplasmic-binding protein	227512	65	514	35.99		*Psi*I	Intergenic
pASC2527	rrn	rRNA operon	rrn					*Psi*I	
pASC2528	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007721	148	511	42.14		*Psi*I	
pASC2529	HI1520	conserved hypothetical protein	1589270	9	355	48.26		*Psi*I	
pASC2531	HI0091	conserved hypothetical protein	98635	150	378	44.62		*Psi*I	
pASC2532	HI0515	DNA-directed RNA polymerase, beta chain (rpoB)	539035	15	1343	39.79		*Psi*I	
pASC2533	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007100	355	511	42.14		*Psi*I	
pASC2534	pASC18	vector	527					*Psi*I	
pASC2535	pASC18	vector	2176					*Psi*I	
pASC2535	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007423	247	511	42.14		*Psi*I	
pASC2536	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007795	123	511	42.14		*Psi*I	
pASC2538	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1008150	5	511	42.14		*Psi*I	
pASC2539	rrn	rRNA operon	rrn					*Psi*I	
pASC2540	HI0661	hemoglobin-binding protein	704911	589	1010	33.07	*161*	*Psi*I	
pASC2542	rrn	rRNA operon	rrn					*Psi*I	
pASC2543	rrn	rRNA operon	rrn					*Psi*I	
pASC2544	pASC18	vector	146					*Psi*I	
pASC2546	HI0216	type I restriction/modification specificity protein (hsdS)	233892	138	385	34.89	*274*	*Psi*I	
pASC2548	rrn	rRNA operon	rrn					*Psi*I	
pASC2549	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007785	126	511	42.14		*Psi*I	
pASC2549	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1008150	5	511	42.14		*Psi*I	
pASC2550	HI1519	H. influenzae predicted coding region HI1519	1588919	28	135	49.88		*Psi*I	
pASC2551	HI0973	H. influenzae predicted coding region HI0973	1030430	51	289	33.56	*242*	*Psi*I	
pASC2552	rrn	rRNA operon	rrn					*Psi*I	
pASC2554	HI0090	conserved hypothetical protein	97027	169	237	38.54	*119*	*Psi*I	Intergenic
pASC2555	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1008080	28	511	42.14		*Psi*I	

pASC2556	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007732	144	511	42.14		*Psi*I	
pASC2557	HI0213	oligopeptide transporter, periplasmic-binding protein	227512	65	514	35.99		*Psi*I	Intergenic
pASC2558	HI1478	transposase (muA)	1561179	33	687	48.23	*18*	*Psi*I	Intergenic
pASC2559	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007967	66	511	42.14		*Psi*I	
pASC2560	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1006740	475	511	42.14		*Psi*I	
pASC2561	HI0213	oligopeptide transporter, periplasmic-binding protein	227542	35	514	35.99		*Psi*I	Intergenic
pASC2562	HI0947	virulence associated protein C (vapC)	1008583	0	132	37.12		*Psi*I	
pASC2563	HI1477	DNA-binding protein (ner)	1561074	55	89	39.7	*123*	*Psi*I	
pASC2564	HI1515	64 kDa virion protein (muN)	1585880	89	455	46.15		*Psi*I	
pASC2564	HI1522	H. influenzae predicted coding region HI1522	1591348	149	623	41.63	*167*	*Psi*I	
pASC2565	HI1478	transposase (muA)	1561179	33	687	48.23	*18*	*Psi*I	Intergenic
pASC2566	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007660	168	511	42.14		*Psi*I	
pASC2566	HI1263	homoserine acetyltransferase (met2)	1340913	108	358	39.85		*Psi*I	
pASC2566	HI1615	phosphoribosylaminoimidazole carboxylase, catalytic subunit (purE)	1683068	34	164	45.12		*Psi*I	
pASC2567	HI0213	oligopeptide transporter, periplasmic-binding protein, putative	227512	65	514	35.99		*Psi*I	Intergenic
pASC2569	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007721	148	511	42.14		*Psi*I	
pASC2569	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007731	144	511	42.14		*Psi*I	
pASC2569	rrn	rRNA operon	rrn					*Psi*I	
pASC2570	HI0213	oligopeptide transporter, periplasmic-binding protein, putative	227512	65	514	35.99		*Psi*I	Intergenic
pASC2571	HI1477	DNA-binding protein (ner)	1561074	55	89	39.7	*123*	*Psi*I	
pASC2572	HI1478	transposase (muA)	1561179	33	687	48.23	*18*	*Psi*I	Intergenic
pASC2574	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007660	168	511	42.14		*Psi*I	
pASC2575	HI0213	oligopeptide transporter, periplasmic-binding protein	227524	53	514	35.99		*Psi*I	Intergenic
pASC2575	HI0947	virulence associated protein C (vapC)	1008583	0	132	37.12		*Psi*I	
pASC2576	HI1478	transposase (muA)	1561179	33	687	48.23	*18*	*Psi*I	Intergenic
pASC2577	rrn	rRNA operon	rrn					*Psi*I	
pASC2578	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007721	148	511	42.14		*Psi*I	
pASC2579	HI0441	orfJ protein	464132	268	281	40.09	*51*	*Psi*I	
pASC2581	rrn	rRNA operon	rrn					*Psi*I	
pASC2582	rrn	rRNA operon	rrn					*Psi*I	
pASC2583	HI1610	tyrosyl tRNA-synthetase (tyrS)	1676603	180	401	39.9	*217*	*Psi*I	
pASC2584	rrn	rRNA operon	rrn					*Psi*I	
pASC2585	rrn	rRNA operon	rrn					*Psi*I	

pASC2586	pASC18	vector	418					*Psi*I	Hinf junct at 587658
pASC2587	rrn	rRNA operon	rrn					*Psi*I	
pASC2589	rrn	rRNA operon	rrn					*Psi*I	
pASC2590	rrn	rRNA operon	rrn					*Psi*I	
pASC2591	HI0715	ATP-dependent Clp protease, ATP-binding subunit (clpX)	763835	62	411	36.09	**87**	*Psi*I	
pASC2592	pASC18	vector	594					*Psi*I	
pASC2592	pASC18	vector	749					*Psi*I	
pASC2593	rrn	rRNA operon	rrn					*Psi*I	
pASC2593	rrn	rRNA operon	rrn					*Psi*I	
pASC2594	HI1514	H. influenzae predicted coding region HI1514	1585505	595	631	48.71		*Psi*I	
pASC2594	HI1516	phage related protein frameshift mutation	1587554	191	419			*Psi*I	
pASC2595	pASC18	vector	587					*Psi*I	
pASC2596	rrn	rRNA operon	rrn					*Psi*I	
pASC2597	HI1042	metB frameshift mutation	1107856	156	616		*93*	*Psi*I	
pASC2599	rrn	rRNA operon	rrn					*Psi*I	
pASC2600	HI0092	H. influenzae predicted coding region HI0092	100197	52	419	43.44		*Psi*I	
pASC2602	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007622	181	511	42.14		*Psi*I	
pASC2603	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007967	66	511	42.14		*Psi*I	
pASC2604	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1008105	20	511	42.14		*Psi*I	
pASC2605	HI0213	oligopeptide transporter, periplasmic-binding protein	227512	65	514	35.99		*Psi*I	Intergenic
pASC2606	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1008095	23	511	42.14		*Psi*I	
pASC2607	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007957	69	511	42.14		*Psi*I	
pASC2608	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1008095	23	511	42.14		*Psi*I	
pASC2609	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007601	188	511	42.14		*Psi*I	
pASC2610	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007990	58	511	42.14		*Psi*I	
pASC2611	rrn	rRNA operon	rrn					*Psi*I	
pASC2613	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007306	286	511	42.14		*Psi*I	
pASC2614	HI1477	DNA-binding protein (ner)	1560877	32	89	39.7		*Psi*I	Intergenic
pASC2615	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007100	355	511	42.14		*Psi*I	
pASC2616	rrn	rRNA operon	rrn					*Psi*I	
pASC2617	HI0213	oligopeptide transporter, periplasmic-binding protein	227542	35	514	35.99		*Psi*I	Intergenic
pASC2618	HI1477	DNA-binding protein (ner)	1561074	55	89	39.7	*123*	*Psi*I	
pASC2619	pASC18	vector	97					*Psi*I	
pASC2619	pASC18	vector	1293					*Psi*I	
pASC2620	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007721	148	511	42.14		*Psi*I	
pASC2621	HI0213	oligopeptide transporter, periplasmic-binding protein	227512	65	514	35.99		*Psi*I	Intergenic

pASC2621	HI1477	DNA-binding protein (ner)	1560910	0	89	39.7	*297*	*Psi*I	
pASC2622	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1008095	23	511	42.14		*Psi*I	
pASC2623	HI1478	transposase (muA)	1561179	33	687	48.23	*18*	*Psi*I	Intergenic
pASC2624	pASC18	vector	1249					*Psi*I	
pASC2625	HI1518	H. influenzae predicted coding region HI1518	1588340	21	182	50.18		*Psi*I	
pASC2626	rrn	rRNA operon	rrn					*Psi*I	
pASC2627	HI0216	type I restriction/modification specificity protein (hsdS)	233892	138	385	34.89	*274*	*Psi*I	
pASC2627	rrn	rRNA operon	rrn					*Psi*I	
pASC2628	HI1477	DNA-binding protein (ner)	1561074	55	89	39.7	*123*	*Psi*I	
pASC2629	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007706	153	511	42.14		*Psi*I	
pASC2630	rrn	rRNA operon	rrn					*Psi*I	
pASC2631	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007093	357	511	42.14		*Psi*I	
pASC2632	HI1477	DNA-binding protein (ner)	1561074	55	89	39.7	*123*	*Psi*I	
pASC2633	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007601	188	511	42.14		*Psi*I	
pASC2633	HI0947	virulence associated protein C (vapC)	1008583	0	132	37.12		*Psi*I	
pASC2634	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1008150	5	511	42.14		*Psi*I	
pASC2635	HI0275	H. influenzae predicted coding region HI0275	310001	124	551	36	*84*	*Psi*I	
pASC2636	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007348	272	511	42.14		*Psi*I	
pASC2637	HI0213	oligopeptide transporter, periplasmic-binding protein	227512	65	514	35.99		*Psi*I	Intergenic
pASC2639	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1006940	408	511	42.14		*Psi*I	
pASC2640	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007912	84	511	42.14		*Psi*I	
pASC2641	pASC18	vector	587					*Psi*I	
pASC2642	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007622	181	511	42.14		*Psi*I	
pASC2643	HI0749	lexA repressor (lexA)	810084	3	209	38.12	*237*	*Psi*I	
pASC2644	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1008061	34	511	42.14		*Psi*I	
pASC2645	rrn	rRNA operon	rrn					*Psi*I	
pASC2647	HI0551	diadenosine-tetraphosphatase (apaH)	572440	241	275	40.36	**71**	*Psi*I	
pASC2648	rrn	rRNA operon	rrn					*Psi*I	
pASC2649	rrn	rRNA operon	rrn					*Psi*I	
pASC2651	rrn	rRNA operon	rrn					*Psi*I	
pASC2652	HI0854	conserved hypothetical protein	901991	175	253	33.73	**24**	*Psi*I	
pASC2653	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007912	84	511	42.14		*Psi*I	
pASC2654	rrn	rRNA operon	rrn					*Psi*I	
pASC2655	pASC18	vector	536					*Psi*I	
pASC2656	rrn	rRNA operon	rrn					*Psi*I	
pASC2657	HI1515	64 kDa virion protein (muN)	1586182	189	455	46.15		*Psi*I	

pASC2657	rrn	rRNA operon	rrn					*Psi*I	
pASC2659	HI1264	DNA gyrase, subunit A (gyrA)	1342986	458	880	41.74	**41**	*Psi*I	
pASC2660	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1008150	5	511	42.14		*Psi*I	
pASC2661	HI1567	TonB-dependent receptor, putative	1634989	93	670	34.88	**130**	*Psi*I	
pASC2662	rrn	rRNA operon	rrn					*Psi*I	
pASC2663	rrn	rRNA operon	rrn					*Psi*I	
pASC2664	rrn	rRNA operon	rrn					*Psi*I	
pASC2665	pASC18	vector	594					*Psi*I	
pASC2666	rrn	rRNA operon	rrn					*Psi*I	
pASC2667	pASC18	vector	270					*Psi*I	
pASC2668	rrn	rRNA operon	rrn					*Psi*I	
pASC2669	HI0897	multidrug resistance protein B (emrB)	953274	297	510	39.67	**104**	*Psi*I	
pASC2670	rrn	rRNA operon	rrn					*Psi*I	
pASC2671	HI1515	64 kDa virion protein (muN)	1585702	29	455	46.15		*Psi*I	
pASC2671	HI1521	conserved hypothetical protein	1590859	155	168	41.07		*Psi*I	
pASC2675	rrn	rRNA operon	rrn					*Psi*I	
pASC2676	pASC18	vector	1192					*Psi*I	
pASC2678	HI1520	conserved hypothetical protein	1589517	92	355	48.26		*Psi*I	
pASC2680	rrn	rRNA operon	rrn					*Psi*I	
pASC2681	rrn	rRNA operon	rrn					*Psi*I	
pASC2682	HI0887	phosphoribosylaminoimidazolecarboxamide formyltransferase (purH)	939933	25	532	42.54	*123*	*Psi*I	
pASC2683	pASC18	vector	1092					*Psi*I	
pASC2684	rrn	rRNA operon	rrn					*Psi*I	
pASC2685	HI1335	cell division protein (ftsH)	1412379	552	635	41.99		*Psi*I	
pASC2686	HI1282	conserved hypothetical protein	1358379	307	141	39.72	*75*	*Psi*I	Intergenic
pASC2687	rrn	rRNA operon	rrn					*Psi*I	
pASC2688	pASC18	vector	527					*Psi*I	
pASC2689	rrn	rRNA operon	rrn					*Psi*I	
pASC2690	rrn	rRNA operon	rrn					*Psi*I	
pASC2691	HI1520	conserved hypothetical protein	1589517	92	355	48.26		*Psi*I	
pASC2691	rrn	rRNA operon	rrn					*Psi*I	
pASC2692	rrn	rRNA operon	rrn					*Psi*I	

Gray shading indicates multiple insertions mapped in the same sequencing reaction

† Distance of Tn5 insertion from insert/vector junction (**bold** = experimental; *italics* = theoretical)

Appendix F – Mapped Tn5 insertions (pASC18MIN libraries)

Allele/Plasmid	CDS/Vector	Protein/Gene	Mutation coordinate	Disrupted codon	Total codons	Percent GC	Distance from junct (<300nt) †	Library	Comments
pASC2702	pASC18MIN	vector	672					*Psi*I	
pASC2702	pASC18MIN	vector	681					*Psi*I	
pASC2703	pASC18MIN	vector	681					*Psi*I	
pASC2704	pASC18MIN	vector	3					*Psi*I	
pASC2705	pASC18MIN	vector	1137					*Psi*I	
pASC2706	pASC18MIN	vector	3					*Psi*I	
pASC2707	pASC18MIN	vector	18					*Psi*I	
pASC2708	pASC18MIN	vector	1263					*Psi*I	
pASC2709	pASC18MIN	vector	530					*Psi*I	
pASC2710	HI0156	malonyl CoA-acyl carrier protein transacylase (fabD)	172543	185	312	41.99		*Psi*I	
pASC2711	pASC18MIN	vector	425					*Psi*I	
pASC2712	pASC18MIN	vector	54					*Psi*I	
pASC2714	pASC18MIN	vector	621					*Psi*I	
pASC2715	pASC18MIN	vector	652					*Psi*I	
pASC2716	HI0156	malonyl CoA-acyl carrier protein transacylase (fabD)	172543	185	312	41.99		*Psi*I	
pASC2718	pASC18MIN	vector	451					*Psi*I	
pASC2720	pASC18MIN	vector	681					*Psi*I	
pASC2723	HI0156	malonyl CoA-acyl carrier protein transacylase (fabD)	172543	185	312	41.99		*Psi*I	
pASC2724	pASC18MIN	vector	993					*Psi*I	
pASC2801	pASC18MIN	vector	75					*Xmn*I	
pASC2802	pASC18MIN	vector	1445					*Xmn*I	
pASC2803	HI0479	ATP synthase F1, subunit beta (atpD)	500694	317	457	40.12		*Xmn*I	
pASC2804	pASC18MIN	vector	219					*Xmn*I	
pASC2804	pASC18MIN	vector	267					*Xmn*I	
pASC2805	HI0738	dihydroxyacid dehydratase (ilvD)	793484	281	612	43.52		*Xmn*I	
pASC2806	pASC18MIN	vector	25					*Xmn*I	

pASC2807	HI0579	elongation factor G (fusA)	598168	560	700	40.67	*47*	*Xmn*I
pASC2808	rrn	rRNA operon	rrn					*Xmn*I
pASC2810	pASC18MIN	vector	1667					*Xmn*I
pASC2811	HI0579	elongation factor G (fusA)	598168	560	700	40.67	*47*	*Xmn*I
pASC2812	rrn	rRNA operon	rrn					*Xmn*I
pASC2814	pASC18MIN	vector	357					*Xmn*I
pASC2814	HI0199	lipid A biosynthesis (kdo)2-(lauroyl)-lipid IVA acyltransferase (msbB)	215126	86	318	40.04		*Xmn*I
pASC2815	HI1429	phosphoribosylaminoimidazole synthetase (purM)	1517846	136	344	44.57		*Xmn*I
pASC2815	HI1430	short chain dehydrogenase/reductase	1518877	147	252	41.67		*Xmn*I
pASC2816	rrn	rRNA operon	rrn					*Xmn*I
pASC2817	pASC18MIN	vector	261					*Xmn*I
pASC2818	HI1356	4-alpha-glucanotransferase (malQ)	1435936	420	699	38.96	*231*	*Xmn*I
pASC2820	HI0579	elongation factor G (fusA)	598168	560	700	40.67	*47*	*Xmn*I
pASC2821	rrn	rRNA operon	rrn					*Xmn*I
pASC2822	HI0196	chorismate synthase (aroC)	212080	21	357	42.39	*100*	*Xmn*I
pASC2823	HI0006	formate dehydrogenase, alpha subunit (fdxG)	6831	391	1029			*Xmn*I
pASC2824	rrn	rRNA operon	rrn					*Xmn*I
pASC2825	rrn	rRNA operon	rrn					*Xmn*I
pASC2826	pASC18MIN	vector	674					*Xmn*I
pASC2827	pASC18MIN	vector	333					*Xmn*I
pASC2827	pASC18MIN	vector	754					*Xmn*I
pASC2828	HI1515	64 kDa virion protein (muN)	1586486	291	455	46.15		*Xmn*I
pASC2829	HI0625	TRK system potassium uptake protein (trkA)	666451	305	458	38.86		*Xmn*I
pASC2830	pASC18MIN	vector	327					*Xmn*I
pASC2831	pASC18MIN	vector	1262					*Xmn*I
pASC2832	pASC18MIN	vector	314					*Xmn*I
pASC2833	HI1501	H. influenzae predicted coding region HI1501	1574457	131	520	48.33		*Xmn*I
pASC2834	pASC18MIN	vector	310					*Xmn*I
pASC2835	pASC18MIN	vector	384					*Xmn*I
pASC2836	rrn	rRNA operon	rrn					*Xmn*I
pASC2837	pASC18MIN	vector	1499					*Xmn*I
pASC2838	pASC18MIN	vector	1285					*Xmn*I
pASC2839	pASC18MIN	vector	1327					*Xmn*I

pASC2840	HI0339	primosomal protein N' (priA)	366749	460	730	41.28	*201*	*Xmn*I
pASC2841	HI0197	penicillin-insensitive murein endopeptidase (mepA)	213446	90	286	43.59		*Xmn*I
pASC2842	pASC18MIN	vector	176					*Xmn*I
pASC2843	HI0909	preprotein translocase SecA subunit (secA)	962434	234	901	40.77		*Xmn*I
pASC2844	pASC18MIN	vector	1025					*Xmn*I
pASC2845	HI0986	2-isopropylmalate synthase (leuA)	1044381	358	531	41.12		*Xmn*I
pASC2847	HI0500	conserved hypothetical protein	517769	314	346	40.66	**89**	*Xmn*I
pASC2848	HI0752	phosphoribosylformylglycinamidine synthase (purL)	815775	936	1320	42.1		*Xmn*I
pASC2849	HI0283	2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase/2-oxoglutarate decarboxylase (menD)	318245	38	568	41.02		*Xmn*I
pASC2850	pASC18MIN	vector	1258					*Xmn*I
pASC2852	pASC18MIN	vector	272					*Xmn*I
pASC2853	pASC18MIN	vector	1187					*Xmn*I
pASC2854	rrn	rRNA operon	rrn					*Xmn*I
pASC2855	pASC18MIN	vector	1429					*Xmn*I
pASC2856	HI1515	64 kDa virion protein (muN)	1586486	291	455	46.15		*Xmn*I
pASC2857	rrn	rRNA operon	rrn					*Xmn*I
pASC2858	pASC18MIN	vector	157					*Xmn*I
pASC2859	HI1515	64 kDa virion protein (muN)	1586486	291	455	46.15		*Xmn*I
pASC2860	HI1515	64 kDa virion protein (muN)	1586486	291	455	46.15		*Xmn*I
pASC2861	pASC18MIN	vector	265					*Xmn*I
pASC2862	pASC18MIN	vector	1180					*Xmn*I
pASC2863	pASC18MIN	vector	1697					*Xmn*I
pASC2864	pASC18MIN	vector	674					*Xmn*I
pASC2865	pASC18MIN	vector	1180					*Xmn*I
pASC2866	HI1612	conserved hypothetical transmembrane protein	1678304	387	464	37.93	*110*	*Xmn*I
pASC2867	rrn	rRNA operon	rrn					*Xmn*I
pASC2868	pASC18MIN	vector	1199					*Xmn*I
pASC2869	pASC18MIN	vector	681					*Xmn*I
pASC2869	HI1252	ABC transporter, ATP-binding protein	1328863	334	556	43.29	*175*	*Xmn*I
pASC2870	pASC18MIN	vector	1332					*Xmn*I
pASC2871	rrn	rRNA operon	rrn					*Xmn*I
pASC2872	HI0213	oligopeptide transp, periplasmic-binding protein	227659	487	514	35.99	*104*	*Xmn*I
pASC2872	rrn		rrn					*Xmn*I

pASC2873	HI0752	phosphoribosylformylglycinamidine synthase (purL)	812682	285	1320	42.1		*Xmn*I	Intergenic
pASC2874	pASC18MIN	vector	681					*Xmn*I	
pASC2875	pASC18MIN	vector	681					*Xmn*I	
pASC2876	HI1515	64 kDa virion protein (muN)	1586486	291	455	46.15		*Xmn*I	
pASC2877	pASC18	vector	139					*Xmn*I	
pASC2877	HI0479	ATP synthase F1, subunit beta (atpD)	500694	317	457	40.12		*Xmn*I	
pASC2877	rrn	rRNA operon	rrn					*Xmn*I	
pASC2878	pASC18MIN	vector	1701					*Xmn*I	
pASC2880	pASC18MIN	vector	317					*Xmn*I	
pASC2881	pASC18MIN	vector	314					*Xmn*I	
pASC2882	HI0339	primosomal protein N' (priA)	366749	460	730	41.28	*201*	*Xmn*I	
pASC2882	HI0771	acetyl-CoA acetyltransferase (atoB)	833540	377	393	40.54	**83**	*Xmn*I	
pASC2883	pASC18MIN	vector	511					*Xmn*I	
pASC2884	HI0283	2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase/2-oxoglutarate decarboxylase (menD)	318245	38	568	41.02		*Xmn*I	
pASC2886	HI1285	type I restriction enzyme (hsdR)	1363627	911	1055	43.13		*Xmn*I	
pASC2887	pASC18MIN	vector	703					*Xmn*I	
pASC2888	rrn	rRNA operon	rrn					*Xmn*I	
pASC2888	rrn	rRNA operon	rrn					*Xmn*I	
pASC2889	pASC18MIN	vector	703					*Xmn*I	
pASC2889	HI0283	2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase/2-oxoglutarate decarboxylase (menD)	318245	38	568	41.02		*Xmn*I	
pASC2890	HI1252	ABC transporter, ATP-binding protein	1328863	334	556	43.29	*175*	*Xmn*I	
pASC2891	HI1514	H. influenzae predicted coding region HI1514	1584978	420	631	48.71		*Xmn*I	
pASC2892	rrn	rRNA operon	rrn					*Xmn*I	
pASC2892	rrn	rRNA operon	rrn					*Xmn*I	
pASC2893	rrn	rRNA operon	rrn					*Xmn*I	
pASC2894	HI0339	primosomal protein N' (priA)	366749	460	730	41.28	*201*	*Xmn*I	
pASC2895	HI1520	conserved hypothetical protein	1589313	24	355	48.26		*Xmn*I	
pASC2896	HI0752	phosphoribosylformylglycinamidine synthase (purL)	815485	839	1320	42.1		*Xmn*I	
pASC2896	HI0752	phosphoribosylformylglycinamidine synthase (purL)	816189	1074	1320	42.1		*Xmn*I	
pASC2897	pASC18MIN	vector	674					*Xmn*I	
pASC2898	rrn	rRNA operon	rrn					*Xmn*I	
pASC2899	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007423	247	511	42.14		*Xmn*I	

pASC2899	rrn	rRNA operon	rrn					*Xmn*I
pASC2900	HI0156	malonyl CoA-acyl carrier protein transacylase (fabD)	172543	185	312	41.99		*Xmn*I
pASC2901	pASC18MIN	vector	1545					*Xmn*I
pASC2902	pASC18MIN	vector	74					*Xmn*I
pASC2903	HI1515	64 kDa virion protein (muN)	1586486	291	455	46.15		*Xmn*I
pASC2904	pASC18MIN	vector	46					*Xmn*I
pASC2904	pASC18MIN	vector	679					*Xmn*I
pASC2905	pASC18MIN	vector	227					*Xmn*I
pASC2906	pASC18MIN	vector	1358					*Xmn*I
pASC2907	HI1515	64 kDa virion protein (muN)	1586486	291	455	46.15		*Xmn*I
pASC2908	pASC18MIN	vector	985					*Xmn*I
pASC2909	pASC18MIN	vector	681					*Xmn*I
pASC2910	pASC18MIN	vector	307					*Xmn*I
pASC2911	HI0946.1	L-2,4-diaminobutyrate decarboxylase	1007433	244	511	42.14		*Xmn*I
pASC2912	HI0751	thiol peroxidase (tpx)	812009	126	165	39.8		*Xmn*I
pASC2912	HI0751	thiol peroxidase (tpx)	812149	79	165	39.8		*Xmn*I
pASC2913	rrn	rRNA operon	rrn					*Xmn*I
pASC2914	HI0156	malonyl CoA-acyl carrier protein transacylase (fabD)	172543	185	312	41.99		*Xmn*I
pASC2915	pASC18MIN	vector	303					*Xmn*I
pASC2916	HI0156	malonyl CoA-acyl carrier protein transacylase (fabD)	172543	185	312	41.99		*Xmn*I
pASC2917	rrn	rRNA operon	rrn					*Xmn*I
pASC2918	rrn	rRNA operon	rrn					*Xmn*I
pASC2919	pASC18MIN	vector	388					*Xmn*I
pASC2920	rrn	rRNA operon	rrn					*Xmn*I
pASC2921	HI0579	elongation factor G (fusA)	598168	560	700	40.67	*47*	*Xmn*I
pASC2922	HI0579	elongation factor G (fusA)	598168	560	700	40.67	*47*	*Xmn*I
pASC2924	HI0440	penicillin-binding protein 1A (ponA)	462367	574	864	41.32	**15**	*Xmn*I
pASC2925	HI0579	elongation factor G (fusA)	598168	560	700	40.67	*47*	*Xmn*I
pASC2926	HI0579	elongation factor G (fusA)	598168	560	700	40.67	*47*	*Xmn*I
pASC2927	HI0649	ATP-dependent DNA helicase (rep)	693650	575	670	40.9	*139*	*Xmn*I
pASC2928	HI0579	elongation factor G (fusA)	598168	560	700	40.67	*47*	*Xmn*I
pASC2929	rrn	rRNA operon	rrn					*Xmn*I
pASC2930	HI0213	oligopeptide trans, periplasmic-binding protein	227659	487	514	35.99	*104*	*Xmn*I

pASC2931	HI0579	elongation factor G (fusA)	598168	560	700	40.67	*47*	*Xmn*I
pASC2932	HI0579	elongation factor G (fusA)	598168	560	700	40.67	*47*	*Xmn*I
pASC2933	HI0579	elongation factor G (fusA)	598168	560	700	40.67	*47*	*Xmn*I
pASC2934	HI0579	elongation factor G (fusA)	598168	560	700	40.67	*47*	*Xmn*I
pASC2935	rrn	rRNA operon	rrn					*Xmn*I
pASC2936	rrn	rRNA operon	rrn					*Xmn*I
pASC2937	pASC18MIN	vector	1337					*Xmn*I
pASC2938	pASC18MIN	vector	213					*Xmn*I
pASC2938	pASC18MIN	vector	220					*Xmn*I
pASC2940	pASC18MIN	vector	672					*Xmn*I
pASC2941	pASC18MIN	vector	18					*Xmn*I
pASC2942	HI1142	cell division protein (ftsA)	1210825	55	425	39.14	*85*	*Xmn*I
pASC2943	HI0579	elongation factor G (fusA)	598168	560	700	40.67	*47*	*Xmn*I

Gray shading indicates multiple insertions mapped in the same sequencing reaction

† Distance of Tn5 insertion from insert/vector junction (**bold** = experimental; *italics* = theoretical)