INVESTIGATION OF THE ROLE OF *comA* IN UPTAKE SIGNAL SEQUENCE RECOGNITION IN *HAEMOPHILUS INFLUENZAE*

by

MELANIE ANNE ALEXANDRA BERTRAND

B.Sc., Simon Fraser University, 1996

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

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

(ZOOLOGY)

THE UNIVERSITY OF BRITISH COLUMBIA

December 2005

Abstract

Haemophilus influenzae is a naturally competent bacterial species that preferentially takes up conspecific DNA. It recognizes this DNA by a nine base pair Uptake Signal Sequence (USS) that is over-represented in its own genome. The mechanism by which this USS is recognized is unknown, though knockout mutations in several genes have been shown to eliminate both binding and uptake of DNA. One of these genes may be involved in binding the USS. I attempted to determine if *comA*, one of the candidate genes, was responsible for the sequence specific binding. This gene was amplified under mutagenic PCR conditions, and then used to transform a hypercompetent strain of *H. influenzae*, using a novobiocin resistant PCR amplified marker sequence to identify transformants. Transformants were then screened for a reduction in transformation frequency to identify possible mutants. *comA* from these lines was then sequenced and compared to the known wild type sequence. None of the mutant lines differed in the sequence of *comA* in comparison to the known wild type, so no conclusions may be drawn concerning the role of *comA*. Mapping these mutations was beyond the scope of this project; however, a possible explanation for the reduced transformation frequency phenotypes is described here. Four isolates retained their phenotypes under competence inducing conditions, indicating a change in binding efficiency. None of the isolates possessed a reversion of the hypercompetent parental genotype.

Table of Contents

List of Tables

List of Figures

List of Terms and Abbreviations

bp	base pair
cAMP	cyclic Adenosine Monophosphate
CFU	Colony Forming Units
CRE	Competence Regulatory Element
CRP	Cyclic AMP Receptor Protien
dGTP	2'-deoxyguanosine 5'-triphosphate
(ds)DNA	(double stranded) Deoxyribonucleic acid
dNTP	Equimolar mix of nucleotides
EMS	Ethyl methanesulfonate
Kan^R / Kan^S	Kanamycin resistant / Kanamycin sensitive
Kb	Kilobase
MIV	'M-four' competence inducing medium
NAD	Nicotinamide Adenine Dinucleotide
Nov^R / Nov^S	Novobiocin resistant / Novobiocin sensitive
PBS	Phospho-buffered saline
PCR	Polymerase Chain Reaction
RNA	Ribonucleic acid
sBHI	Supplemented brain-heart infusion
TF	Transformation Frequency
USS	Uptake Signal Sequence

1. Introduction

1.1 Introduction

Natural competence is the ability of bacteria cells to import exogenous DNA. Transformation, the change in genotype by recombination with this DNA, may then occur. This provides a method by which a cell could gain new pathogenesis genes, resistance to antibiotics, or variations in capsular characteristics [1, 2]. *Haemophilus influenzae* is a particularly interesting naturally competent bacteria species because it specifically binds and takes up DNA from its own species. This species-specific uptake is accomplished by recognizing a short DNA sequence called an Uptake Signal Sequence (USS) [3, 4]. Little is known about the mechanism by which this USS is recognized, despite many experimental attempts. In this thesis I describe an attempt to characterize ComA, a candidate USS binding protein, by directed mutagenesis, and investigations of the resulting mutant strains. Identification of the components of this mechanism and an understanding of the manner in which they function will assist in gaining a more complete understanding of natural competence.

H. influenzae is an important human pathogen, causing pneumonia and infections of the sinuses and inner ears. The young, elderly, and immunocompromised are particularly vulnerable [5]. Serotype b strains can cause infant meningitis [6], and in 1990 an effective vaccine against these strains was developed. Other serotypes are unaffected by this vaccine and are still clinically important causes of diseases, as are strains lacking a capsule (nontypeable) [5]. Antibiotic resistance has not been a serious problem, but is

increasing (for example, [7]). Approximately 75% of healthy adults and children have *H. influenzae* commensally inhabiting the mucous membranes of their respiratory tracts, without causing disease [8].

The laboratory strain KW20 is a noncapsular ('rough') strain, isolated from the nasopharynx of a healthy male individual. The genome sequence of this strain was completed in 1995 [9]. This resource allows both comparison of *H. influenzae* to other competent bacteria and detailed genetic study of competence.

1.2 Natural competence

The ability to become competent is widely distributed across the major divisions of Bacteria. However, it is irregular in occurrence; often a competent species will be the closest phylogenetic relative to bacteria that are not competent in the laboratory. Well-studied competent bacteria include the gram positive *Bacillus subtilis* and *Streptococcus pneumoniae*, and the gram negative *Neisseria gonorrhoeae* and *H. influenzae* [10 , 11, 12]. Some components of the uptake machinery, such as the Type IV pilins, are shared between these bacteria, although gram positive bacteria have a single membrane and a thick outer cell wall across which DNA must be transported, and gram negative bacteria have two membranes with a thin wall between them [13, 14]. *Helicobacter pylori* is an exception, having an independently evolved uptake mechanism [15].

1.2.1 *Haemophilus/Neisseria* model

The general process of DNA uptake by gram negative bacteria is illustrated in Figure 1. DNA is bound to the outer membrane, shown as point A. With one known exception,

Acinetobacter [16], gram negative bacteria preferentially bind conspecific DNA, though in the case of *Pseudomonas stutzeri* this specificity is less stringent [10]. Both *Neisseria* and *Haemophilus* accomplish homospecific binding by recognizing a short (10 and 9 base pairs respectively [3, 4]) sequence, called the Uptake Signal Sequence (USS) in *Haemophilus*. The genome of *H. influenzae* is enriched for this sequence, with 1471 occurrences of the exact sequence. Only eight occurrences are predicted in a random sequence of similar GC richness and length [17, 18]. The mechanism by which the cell recognizes the USS and binds to this sequence is unknown; several attempts to identify a USS binding protein have failed.

Binding and transport of DNA across the outer membrane ('uptake') may or may not be integrated as a single process [19], though Barouki and Smith [20] detail a successive-wash method by which to differentiate reversible binding and uptake across the outer membrane. DNA is then transported across the inner membrane ('translocation') [21], shown at point B of Figure 1. Circular and hair-pin DNAs are not translocated, indicating that a free end of DNA is required for the molecule to cross the inner membrane [22]. The free end is thought to be threaded through a complex in the inner membrane, and immediately before, during or after this process the strand entering 5' end first is degraded. The other strand [23] is slowly degraded, and is also available for homologous recombination (point C) during this process. The nucleotides resulting from degradation of incoming DNA are used primarily for DNA synthesis [24].

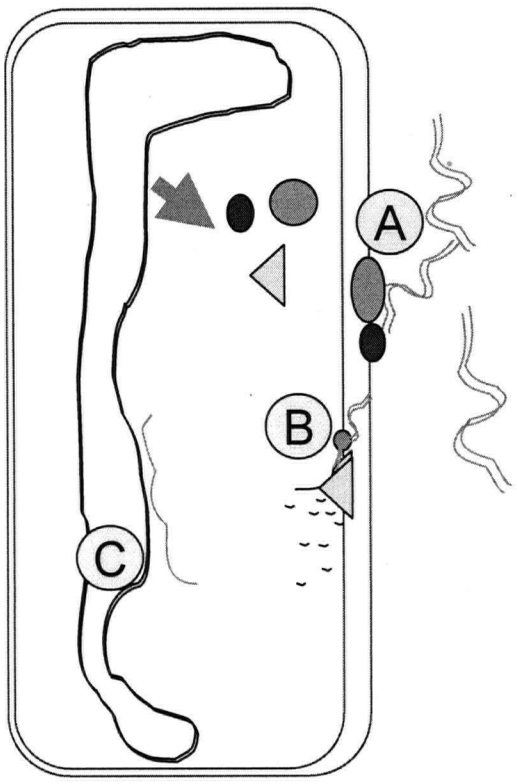


Figure 1: *Neisseria/Haemophilus* model of natural competence.

(**A**) USS containing DNA is bound and transported across the outer membrane. (**B**) Transportation across the inner membrane. (**C**) Homologous recombination may occur between the incoming DNA and the genome.

Transformation efficiency in *H. influenzae* depends strongly on the environment in which it is studied. When cultured in rich, hemin and NAD supplemented liquid medium, KW20 has a transformation frequency ranging from $1x10^{-8}$ in early log growth to $1x10^{-4}$ at late log growth. To induce maximal transformation frequencies (1 to $2x10^{-2}$), cells are transferred to a defined starvation media, as described in section 2.2.3 [25]. Cells within colonies grown on agar plates have transformation frequencies of approximately $1x10^{-5}$ [26].

1.3 Evolution of Competence

1.3.1 Risks and costs

Studies of evolution often focus on benefits of a trait, but need to consider both benefits and costs. Natural competence may be a dangerous process for the cell. Recombination with foreign DNA is likely to introduce mutations. Although the majority of mutations are expected to be selectively neutral, the DNA accessible to transformable cells is from dead cells in the environment. Selectively killed cells can be considered to possess traits non-adaptive to the environment, and the alleles are predicted to contribute an additional cost to transformation [27]. Furthermore, the ability of the cell to be transformed would be lost if the acquired DNA contains a mutation in genes responsible for transformation. The uptake of phage or prophage DNA could be lethal [28]. Uptake of non-homologous DNA triggers the SOS response [29], which has been shown to kill *H. influenzae* cells by inducing a prophage contained within the genome. In addition to these risks, there is a

cost associated with developing competence: the competent cell must expend the energy and resources needed to synthesize the uptake machinery and transport DNA across two membranes.

1.3.2 Benefits

Studies of the regulation of expression of genes encoding proteins involved in competence can be used to help us understand the benefits of natural competence by indicating the circumstances where competence may be advantageous. In *H. influenzae* the characteristics of this regulation best support a nutritional role for DNA uptake. Competence develops under limiting nutritional conditions, and specifically purine depletion [30]. DNA may be broken down into nitrogen and sugars, and it may be used as building blocks for DNA or RNA synthesis. Synthesis of nucleotides is an energetically expensive cellular process [31, 32], and DNA is abundant in the mucosal environments that *Haemophilus* and many other competent bacteria inhabit [10]. This suggests that uptake of DNA is initiated by a nutritional cue [33, 34]. Competence genes are expressed in direct response to an increase in cAMP levels. A sequence in their promoter regions, called the competence regulatory element (CRE), resembles the binding consensus sequence recognized by the cAMP receptor protein (CRP) (see [35] for more detailed information).

Other benefits of transformation have been proposed. One benefit may be the gain of new alleles by recombination, and new combinations of traits. Negating this benefit is that a beneficial combination may then be disrupted by the process by which it was

created [27, 36]. Another possible explanation is that DNA uptake facilitates repair of damaged DNA [37] However, DNA uptake is not triggered by DNA damage by mitomycin C or UV damage to cells' DNA [38]. The lack of relationship between DNA damage and development of competence indicates that the benefit of competence does not lie with DNA repair.

1.5 Genetics of Competence

Although my research focuses on the ability of *H. influenzae* to recognize the USS, this cannot be completely studied in isolation from the remainder of the system. The competence regulon and its members are thoroughly reviewed in [39]. My research was completed prior to the microarray analysis that characterized the regulon, and so this useful information was not available during my experiments.

1.5.1 Determining the roles of competence genes

The phenotype of mutant strains can indicate the role of competence genes. The easiest screen for competence mutants identifies cells incapable of transformation. Further tests, using radio-labeled DNA, can discern defects in the ability to take up DNA and to bind DNA. In *Neisseria*, these processes are mechanistically separable [40], but this may not be true in *H. influenzae*. Alternatively, mutations that have regulatory effects may prevent DNA uptake. In this case, other competence genes will not be induced normally.

Further deductions can be based on gene and protein sequences, which allow predictions of function based on homologous genes in other bacteria. Computer programs such as ProDom compare amino acid sequences against a database of motifs characteristic of protein functions or classes [41, 42]. The presence of a functional motif may indicate the general role of a protein. Other programs (such as SignalP) use amino acid sequences to predict the location of a protein within a cell [43].

The availability of other genomes allows comparisons of *H. influenzae* to other competent species. At the time of my research, the genomes of *Pasteurella multocida, Actinobacillus actinomycetemcomitans , N. meningitidis* and *N. gonorrhoeae* were available in various stages of completion. *P. multocida* is not transformable in the lab [11]. *A. actinomycetemcomitans* has been shown to be competent, and also possesses a core USS identical to that of *H. influenzae*, bounded by similar flanking sequences [44]. However, few mutant studies on this organism have been done. *N. gonorrhoeae* is constitutively competent, and as previously mentioned has an uptake sequence with no similarity to the *H. influenzae* USS. Competence gene homologues have been found in both *A. actinomycetemcomitans* and *N. gonorrhoeae* (for example, [45]), and it is expected that the protein responsible for USS recognition in *A. actinomycetemcomitans* will be very similar to the USS recognition protein in *H. influenzae.*

1.5.2 Genes known to affect Competence

Redfield *et al.* [39] have identified a suite of 25 genes that form a competence regulon. These genes are characterized by the presence of an upstream competence regulatory

element (CRE), and a dramatic increase of expression associated with the development of competence. The 22bp CRE is closely related to the consensus binding sequence of CRP. Expression of the competence regulon genes depends strongly on cAMP concentration, and requires both CRP and Sxy, a protein thought to direct binding of CRP to the CRE [34, 35, 46].

Genes outside of the competence regulon also affect competence, affecting the cell's ability to take up or be transformed by DNA by acting in cellular functions such as transcription and translation, protein export, and cell wall synthesis. In particular, the peptidoglycan biosynthesis gene *murE* is interesting, as specific point mutations in this gene cause a constitutively high level of competence. The mechanism of this effect is unknown [47].

1.5.3 Binding and uptake

Binding and uptake of DNA is the specific step that I investigated. Little is known about the process of binding DNA and transporting it across the outer membrane in *H. influenzae*, but several genes have been identified as functioning at this level. Because a mutation that prevents DNA binding and uptake may do so by physically interfering with binding or uptake, or by preventing the expression of the competence regulon, a binding and uptake candidate gene must be shown to not have a regulatory effect.

Three of the genes known to be needed for DNA binding and uptake (*comA, comC* and *comE)* are members of the *comABCDEF* operon (*comA-F*), illustrated in Figure 2

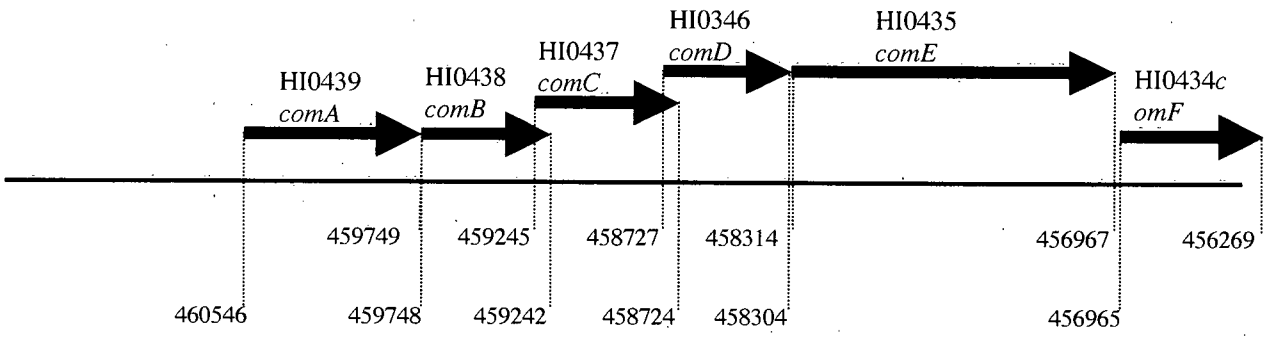


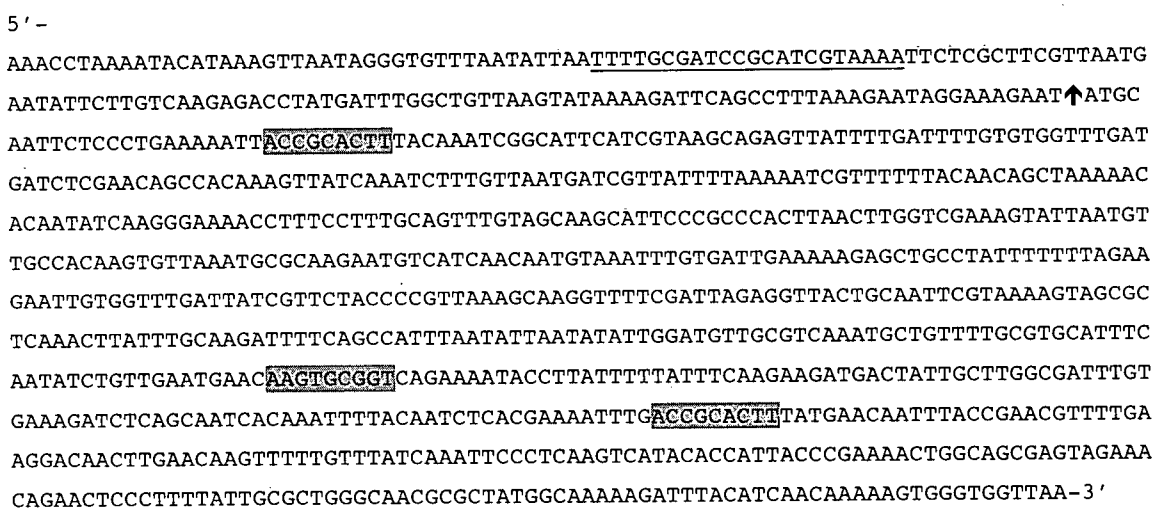
Figure 2: The *H. influenzae comA-F* operon. Arrow placement indicates reading frame. Numbers indicate relative bp position (coordinates) on the chromosome

(http://www.tigr.org/tigr-scripts/CMR2/GenomePage3.spl?database=ghi).

[48, 49]. Mutations of these three genes prevent uptake of DNA in cells that should otherwise be competent. This operon has a CRE located upstream of *comA*, and the operon is induced in competent cells [39]. As seen for other CRE regulon genes, knockouts in the genes encoding CRP (*crp),* Sxy *(sxy),* and Adenylate cyclase (*cya)* abolish expression of the *comA-F* operon [26, 35, 46, 50, 51].

Characteristics of *comA, comC,* and *comE* further support their involvement in DNA uptake. ComA (HI0435) has homologues in the competent bacteria *P. multocida* and *A. actinomycetemcomitans*, but its function in these bacteria is unknown. It lacks any recognizable sequence motif and is predicted to be a globular protein. Features of the nucleotide sequence of *comA* and its upstream region, and of the amino acid sequence, are shown in Figure 3. ComC (HI0437) has a weak homologue of unknown function in *P. multocida.* This protein is predicted to contain a transmembrane helix in its N-terminal region, but lacks the cleavage sequence characteristic of proteins destined for the outer membrane. Lastly, *comE* (HI0435) has homologues in diverse groups of bacteria. It is expected to be an outer membrane protein, and it contains a sequence motif characteristic of the PilQ family proteins, part of the Type IV secretion system. In *Neisseria* PilQ is known to be essential for transformation [52]. PilQ proteins form dodecameric pores in outer membranes, with diameters of approximately 53 to 60 angstroms. As double-stranded DNA (dsDNA) has a diameter of approximately 20 angstroms, this means that a homologous ComE pore should be large enough to transport the dsDNA taken up by *H. influenzae*.

A.



B.

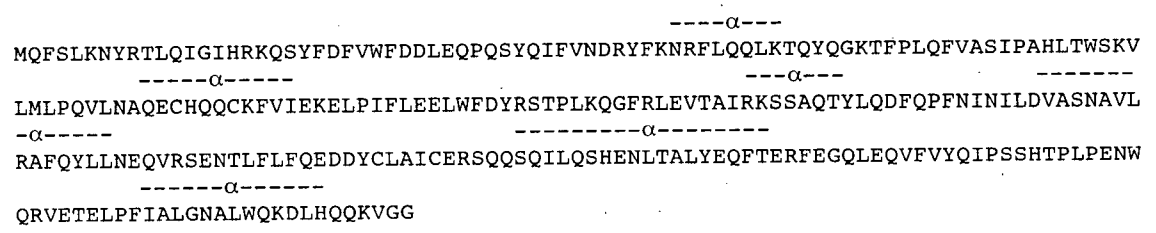


Figure 3: Features of the nucleotide sequence and amino acid sequence of *comA*. A) nucleotide sequence of coordinates 459749 to 460700 of the *H.influenzae* genome. Start codon of *comA* lies immediately to the left of the arrow. CRE binding site is underlined [35]. USSs are shaded grey. B) Amino acid sequence of comA. α helixes shown where predicted by PROF [53] with a reliability index of greater than 5 (on a scale of 0-9) for all involved amino acids.

No mutations of *comB* (HI0438) and *comD* (HI0436) have been identified. These genes are anticipated to be involved in competence due to their location in an operon with known competence genes, but neither possesses any recognized sequence motifs. *ComF* (HI0435) mutants have normal DNA uptake across the outer membrane, but DNA translocation across the inner membrane is prevented. ComF is therefore not thought to be involved with DNA binding and uptake [54, 55].

It should be noted that polarity may be an issue within the *comA-F* operon; if so the phenotype of *comA* knockouts might not be due to the role that ComA plays. This also applies to *comC* knockouts. These mutations were caused by mini Tn*10* insertions. This transposon contains transcription-terminating sequences that prevent transcription of downstream genes (see Brewster and Siehnel [56, 57] for examples where transcription is not terminated). It also contains stop codons. Normally, within an operon, the start codon of a gene is located near the stop codon of the previous gene. A greater distance between these codons reduces the chance that the later gene is translated [58]. If *comD* is involved in binding and uptake of DNA, a knockout mutation in *comA* or *comC* that prevents transcription or translation of *comD* will have a loss of binding and uptake phenotype. Polarity is not an issue with the *comE* knockout. *ComF* is the only gene downstream from *comE*, and its function is in translocation of DNA. If the phenotype of the *comE* knockout was due to polarity, this mutant would have a translocation dysfunction, not a loss of binding.

Binding and uptake also depends on genes outside the *comA-F* operon. Recently, the operon HI0938-41 was identified as belonging to the competence regulon. It has a CRE and is strongly induced during competence. The genes of the operon have homologues in a wide range of bacteria. The first gene of the operon, *pulG*, is a putative fimbrial-like protein, and the second, *pulJ*, is a putative chaperone. Mutants of these genes are unable to bind or take up DNA [59, 60].

comE1 (HI1008) may be involved with DNA binding [12]. This protein has homologues in *Neisseria* and *S. pneumoniae*, and a partial homologue in *B. subtilis* that is essential for DNA uptake [61]. A knockout of this gene reduces transformation in *H. influenzae* by 25-fold [62]. These knockouts retain the ability to differentiate between *H. influenzae* and *E. coli* DNA, indicating that this gene is not involved in recognition of the USS [59].

A complex mutation that is a combination of a cassette insertion in *orfJ* (HI0421) and a deletion of a large part of the downstream gene HI0422 abolishes both binding and uptake of DNA [63]. A knockout of the *orfJ* homologue in *E.coli* is unable to survive with DNA as its sole carbon source, suggesting a conserved role in DNA metabolism [64]. However, *orfJ* is not induced during development of competence [39], and is therefore not likely to be directly involved in binding and uptake.

As knockouts of *comA, comC, comE, orfJ, pulG, pulJ* and *comE1* all prevent binding and uptake of DNA, they are expected to play some role in the mechanism of DNA uptake. This role may or may not be direct; they may play a part in regulation of competence or

assembly of a binding and uptake complex rather than participate directly in binding or uptake.

1.4 The study of genes by mutation

I am interested in characterizing the mechanism by which *H. influenzae* recognizes the USS. As described below, the mutants we currently have that are deficient in DNA binding offer no information concerning how they function. Therefore, my goal was to create more informative mutants.

1.4.1 Transposon mutagenesis

The strains with mutations affecting DNA binding (*comA, comC, comE, orfJ, pulG, pulJ* and *comE1)* were created by transposon or cassette mutagenesis. Transposons and cassettes are used to interrupt genes, and are relatively easy to work with as they contain a selectable antibiotic resistance marker. They contain stop codons and may contain transcription terminating sequences, and often create frameshift mutations. The loss-of-function phenotype may indicate the process the protein is involved in, but usually cannot show how an individual protein acts, as this may be confounded by loss of function of associated proteins. This method would therefore be inappropriate for my research.

1.4.2 Point mutations

Mutations that alter the protein in more subtle ways can demonstrate more of the protein's characteristics. Point mutations may substitute amino acids, reducing the

efficiency of a protein's function, slightly altering binding sites, or may create stop codons. This allows protein interactions or specific function of a protein to be more easily deduced and investigated.

The classic method of creating point mutations is to use chemical mutagens such as ethyl methanesulfonate (EMS). Unlike transposons and cassettes, this method does not introduce a selectable marker. Instead, phenotypic screens are used to identify possible mutants. Such screening for transformation defects is very laborious. As the mutation may be anywhere in the genome, it can be time and labour intensive to map these mutations.

To create point mutations at specific locations, altering function of identified functional regions of the protein, the Polymerase Chain Reaction (PCR) can be used. In the classic site-directed mutagenesis technique, specific mutations can be introduced, targeting known functional regions of the protein. *ComA* has no functional motif that would be useful for a site-directed method of mutagenesis. In this case, a more random approach to mutagenesis within the gene would be more useful.

The error rate of PCR can be used to create randomly placed mutations within a targeted sequence. These error rates range from 0.27 to greater than one point mutation per Kb of amplified DNA in an ordinary 30 cycle reaction, depending on the polymerase used and the conditions of the reaction [65-67]. Addition of manganese to the reaction, or shifting the normally equimolar dNTP concentrations to a higher relative dGTP concentration, further increases the error rate [68, 69]. Combined, these two methods can promote error

rates ranging from 2.0 to 8.1 point mutations per Kb at random locations within the amplified DNA [70]. Clontech supplies a kit that allows fine-tuning of this mutation rate by varying the concentrations of dGTP and manganese. The types of substitutions created varies, depending on the mutagenic conditions. A mutation rate of approximately 4.8 point mutations per Kb is suggested as being appropriate for studies of phenotype alterations, as opposed to identification of the function of individual amino acids [71].

Transformation of competent cells can be used to introduce an assortment of mutations to a population of bacteria by chromosomal recombination. The mutagenized population can then be screened for changes in phenotype [72], and the region can then be sequenced to identify the exact location of the mutations. I used mutagenic PCR to cause mutations in *comA,* then transformed the hypercompetent strain RR804 [47] with the resulting products. The colonies were then screened for transformation defects.

1.6 The Evolution of the USS

1.6.1 The USS

H. influenzae cells specifically bind and take up DNA containing the nine base pair sequence 5'-AAGTGCGGT-3' [4]. As mentioned earlier, the *H. influenzae* genome is highly enriched in this sequence and in sequences differing by a single base. Flanking sequences are also conserved: 5' to the sequence, there is a conserved A, and the 3' consensus sequence is 19 bases rich in A/T content. USSs occur with equal frequency in both strands of the genome, in a non-random distribution [73].

There are two issues to consider when discussing the evolution of the USS. First, which came first – the over-representation of the sequence that later became the USS, or specificity for the uptake of USS that caused the development of the over-representation? The second issue is the source of the benefit of uptake specificity. Here, I use this framework to present two models for the evolution of the USS. Both models provide testable predictions of the consequences of removal of the selectivity in binding, and are the basis of my research.

1.6.2 Model I: Benefit from binding

The first model of USS evolution is that the sequence-specific uptake system developed as a mechanism to increase the efficiency of DNA uptake as a carbon source. The interaction between sequence-specific proteins on the cell surface and their preferred DNA sequence may be stronger and last longer than a non-specific complex. This would increase the likelihood that a bound molecule of DNA would remain bound long enough for uptake. The benefit is then in the frequency of successful uptake events.

In this model, the existence of a recognition sequence in the genome prior to development of a recognition sequence is not necessary but is not precluded. Any preferred binding sequence could become over-represented in the genome. Incoming DNA carrying this USS would occasionally be incorporated into the cell's genome by homologous

recombination, gradually increasing the frequency of the USS. Since I began my research, strong support for this model has been presented [44, 74].

1.6.3 Model II: Benefit from exclusion of foreign DNA

The alternative model is that USS recognition developed due to an advantage of excluding potentially harmful foreign DNA. In this case, the USS must have been already over-represented in the genome in order to provide this means of identifying conspecific DNA.

A sequence is unlikely to become sufficiently overrepresented to drive this selection by chance. What role could the sequence play in the cell? It has been suggested that the USS could function as a transcription terminating sequence as proposed for *Neisseria* [4], or as a Chi sequence, but these possibilities can be refuted by analysis of the locations of USSs in comparison to the locations of sequences involved in the above functions [18, 75]. However, there may be some other as-yet unknown structural function for the USS, which would support the pre-existing sequence bias model.

1.6.4 Prediction of USS-recognition deficient mutant phenotype

Based on what is known of DNA uptake, predictions of the phenotype we would expect in USS recognition mutant cells can be developed for each of the two models. If the USS recognition serves to increase the efficiency of DNA uptake, mutation of the recognition site should severely reduce uptake. Due to the loss of USS recognition, DNA from *H.*

influenzae and *E. coli* should be taken up at equal but low rates. The predictions of the exclusion of foreign DNA model are less clear. A mutation in the USS recognition site may cause the recognition protein to act as a barrier to all DNA, abolishing uptake. Alternatively, removal of the sequence recognition ability of the protein may increase uptake of foreign DNA to levels similar to *H. influenzae* DNA uptake. Thus, providing that mutation of the USS recognition protein does not abolish uptake, the uptake rate of *E. coli* DNA relative to that of *H. influenzae* DNA may offer support for one model over the other.

Some further predictions can made about the unidentified USS recognition protein. The protein must be located on the outer face of the outer membrane. The protein that recognizes the USS may be an integral component of the uptake mechanism, or it may be peripheral. We know that the protein must associate with other members of the uptake apparatus, and that the protein must bind DNA on its own, in a multimer, or with other proteins. Some, but not all DNA-recognizing proteins contain amino acid motifs, such as a helix-turn-helix pattern or a leucine zipper to name two of a large number of DNA-binding motifs. None of the proteins identified as being involved in the mechanism of competence have such a sequence.

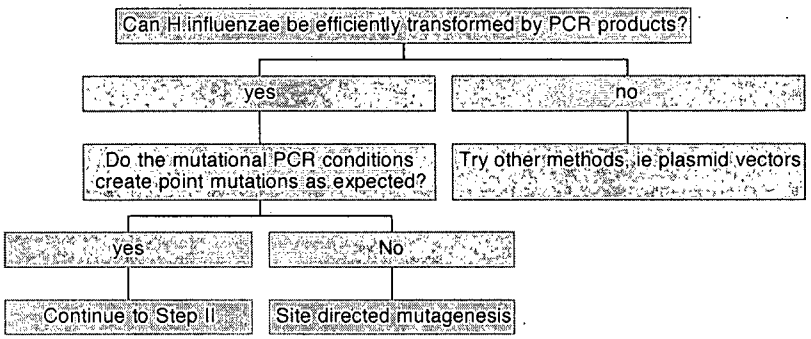
1.7 Experimental Strategy

comA is one of the genes involved in DNA uptake, and thus may be involved in the specific binding of the USS. It may play a role in assembly of the binding and uptake mechanism, it may function as a structural part of the mechanism, or it may directly bind

or transport DNA. To answer the question, "Do point mutations in *comA* alter the ability of *H. influenzae* to recognize the USS," it was first established that ComA is not involved in the regulation of competence. I then followed the two-step procedure outlined in Figure 4. After testing the mutation and transformation protocols in Step I, I used mutagenic PCR to create a population of *comA* sequences with point mutations, in Step II. RR804, a hypercompetent strain, was then transformed with the PCR products. This strain was chosen as the parental type because high competence levels facilitated the mutant screen [47]. Transformants were screened for decreased transformation frequency with respect to the hypercompetent parental strain, then *comA* was sequenced from all isolates with consistently low transformation frequencies.

Do point mutations in *comA* alter the ability to recognize the USS?

Step I: Establishing mutational protocol



Step II: Mutation of *comA* and screening for mutants

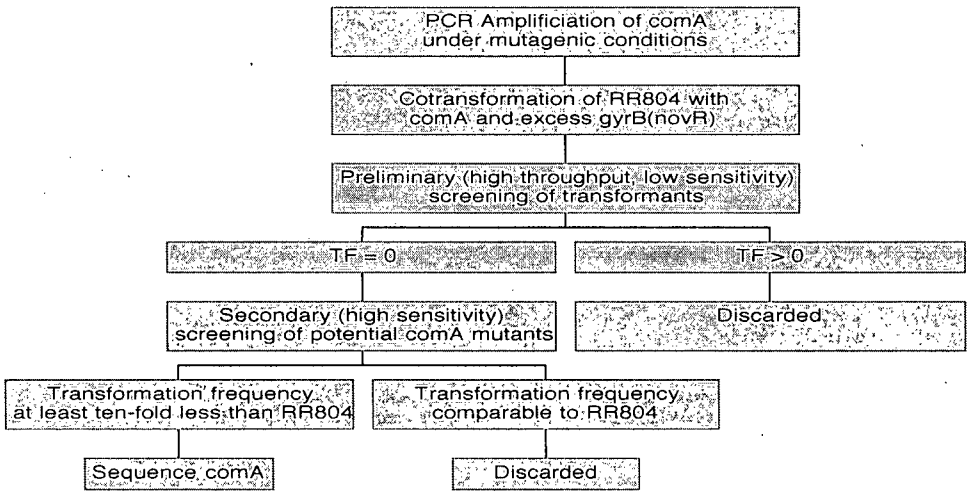


Figure 4: Organization of the experimental strategy.

2. General Materials and Methods

2.1 Culture Techniques and Strains

The strains used in this study are listed in Table 1. All strains are descended from the original KW20 strain [76] of *Haemophilus influenzae* rough serotype d [77]. Gene numbers (HI#) and names are those assigned by Fleischmann *et al.* [9] and The Institute for Genomic Research (TIGR).

Unless otherwise specified, strains were streaked to yield colonies on appropriate antibiotic selective plates and single colonies were chosen for analysis, or strains were cultured from a 1 ml sample frozen in early log growth; these originated from a single colony.

Strains were cultured at 37°C in brain-heart infusion (BHI) (Difco) medium supplemented with hemin (10 μg/ml, Sigma) and NAD (2 μg/ml, Boerhinger Mannheim) (sBHI) [78]. Microbiological quality agar (12 g/l, Marine Bioproducts) was added for solid media. Plates older than 24 h were supplemented with additional hemin (50 μl of a 1 mg/ml solution) and were not used for transformation frequency experiments. Antibiotics were added to liquid media or to melted agar media to give concentrations as follows, unless otherwise stated: novobiocin, 2.5 μg/ml; kanamycin 7.0 μg/ml; streptomycin, 250 μg/ml; chloramphenicol 1 μg/ml.

Table 1: *H. influenzae* strains used in this study.

Strain name	Relevant genotype	Reference
KW20	Wild type	[76]
MAP7	Cm^R, Nov^R, Kan^R	[79]
RR804	*murE*$^-$	[47]
JG48	*comA::tn*10 Kan^R	[48]
Jg7	*comC::tn*10 Kan^R	[48]
Jg1	*comE ::tn*10 Kan^R	[48]
RR871	*rec-2::lacZ* - Cm^R	[80]
RR885	*cya*$^-$, *rec-2::lacZ*, Cm^R	[80]

Strains were incubated in a 37°C incubator, either stationary or, if mild aeration for small amounts of liquid media was required, in 10 ml culture tubes on a tissue culture roller at approximately 60 rpm. If vigorous aeration was required, cultures were grown in an Erlenmyer flask of at least five times the culture volume, shaking at 200 rpm in a heated water bath. Density of cultures was measured using a spectrophotometer, using the absorbance of the culture at 600 nm (OD600). For plating, serial dilutions of a culture were made in dilution solution. This consisted of 1x Phosphate Buffered Saline (PBS) solution (components listed in Table 2) supplemented with approximately 4% BHI. Stocks were frozen for storage at -70°C in 15% glycerol, in log growth phase for storage of stock strains, or in MIV medium (components listed in Table 3), ready for transformation. All components of media and solutions were sterilized by autoclaving or by filtration, except hemin. Hemin suspension was prepared in 4% triethanolamine, then incubated at 65°C for 30 minutes.

2.2 Transformation

Several different techniques were used in transformation of cells. The general procedures follow. Methods have been described in [47, 78, 81, 82].

Table 2: Phosphate-buffered saline.

0.3 g KH_2PO_4
1.1 g $Na_2HPO_4 \cdot 7H_2O$
8.5 g NaCl
dH_2O to 1 L

Table 3: MIV medium: 1 ml of each of solutions 22, 23, 24 and 40 is added to 100 ml of solution 21.

	Amount
Solution 21	
Distilled water	850 ml
L-Aspartic acid	4.0 g
L-Glutamic acid	0.2 g
Furmaric acid	1.0 g
NaCl	4.7 g
Tween 80	0.2 ml
Potassium phosphate (dibasic)	0.87 g
Potassium phosphate (monobasic)	0.67 g
Solution 22	
L-Cystine	0.04 g
L-Tyrosine	0.1 g
L-Citruline	0.06 g
L-Phenylalanine	0.2 g
L-Serine	0.3 g
L-Alanine	0.2 g
Solution 23	
CaCl2	0.1 M
Solution 24	
MgSO4	0.1 M
Solution 40	
5% w/v solution of vitamin-free casamino acides (Difco)	

2.2.1 Hypercompetent Strains

RR804 has a point mutation in *murE*, a peptidoglycan synthesis gene. RR804 develops high competence levels, comparable to induced wild type, during log growth phase (*i.e.* without induction) [47]. For transformation of RR804, eight to ten colonies were picked one day after plating, and suspended together in 5 ml sBHI. The test tube was rolled for fifteen minutes at 37°C then aliquots were taken for transformation with DNA concentrations as specified for each experiment.

2.2.2 Colony Competence

To determine the average competence of all viable cells in a colony, colonies were picked 24 to 36 hours after plating. Each colony was suspended in 5 ml sBHI containing 1 μg/ml MAP7 chromosomal DNA obtained from lab stocks. These were incubated, standing, for 15 min at 37°C, before plating to determine transformation frequency [26].

2.2.3 Induction of Competence Using MIV

15 ml of culture at log growth in sBHI (OD600 = 0.2 to 0.3) were collected onto a 0.2 μm filter using a Nalgene analytical filter funnel, and washed with 15 ml of MIV medium. The filter was then placed in 15 ml of MIV, and shaken for 100 minutes at 100 rpm, 37°C [82]. 10 to 15 ml of culture were filtered, and equal volumes were used for washing and resuspension of cells.

2.2.4 Plating to Determine Transformation Frequency

Once a culture was incubated with DNA containing an antibiotic resistance gene, it was necessary to determine what fraction of cells were transformed. To accomplish this, the culture was treated for ten minutes with DNase I sufficient to digest all DNA used in the transformation (10 μl of 100 μg/μl Dnase I in a 1 ml transformation). Dilutions of the culture were plated on antibiotic-containing sBHI plates and on plain sBHI plates. The number of colonies growing on each type of plate was used, with the dilution factor and the volume plated, to determine the number of colony forming units (CFU) per ml in the culture, and the number of resistant CFU per ml in the culture. The frequency of resistance in the sample was the transformation frequency. With all transformations, a no-DNA control was included.

2.3 DNA Techniques

2.3.1 Chromosomal DNA

Chromosomal DNA was obtained from lab stocks or isolated using techniques described in [82]. Distilled water was used to dilute DNA to appropriate concentrations. The absorbance of DNA at 260 nm, with a conversion factor of 50 μg/ml, was used to determine concentration.

2.3.2 PCR from Chromosomal DNA and Colonies

Amplification of DNA using PCR was used to confirm DNA fragment size, obtain copies of genes, and mutate sequences. Taq DNA polymerase, buffer, and nucleotides were

supplied by Clontech, as was the mutagenic PCR kit Diversify™. Primers were designed using the computer program Genetool [83], and then checked against 30 Kb *H. influenzae* genomic sequences for possible non-target amplification using the computer program Amplify 1.2 [84]. Reactions were carried out in an MJ Research minicycler. The suppliers' protocols were followed for the reaction, using annealing temperatures specific to the primers.

Template for PCR was either chromosomal DNA (2 ng/μl) or cells from a single colony. When a colony was used, a 200 μl pipet tip was touched to the colony to obtain a sample of cells, then immersed in the PCR reagent mix. To lyse the cells, the reactions were held at 94°C for ten minutes prior to normal PCR temperature cycling. A no-template control was run with each reaction. The mutagenesis kit's controls were included each time.

Reaction products were visualized using agarose gel electrophoresis. Unless otherwise noted, 2% agarose gels were used when the product was expected to be less than 1 Kb, otherwise 1% agarose gels were used. Ethidium bromide (0.2 μg/ml) was used to stain the gel, and then the fragments were trans-illuminated with UV light. DNA fragment sizes and concentrations were estimated by comparison to the appropriate sized ladder, using either 1 Kb (GibcoBRL) or 100 bp (Roche) ladders.

2.3.3 Restriction Digests

Digestion of PCR products with restriction enzymes was done according to the manufacturers' specifications. Incubation at the appropriate temperature was done in a 37°C air incubator, a heating block set to 68°C, or using the minicycler.

2.3.4 DNA Sequencing

The PCR amplified DNA sequence of interest was obtained using colonies less than 24 hours of age. To ensure purity of the product, bands were cut from agarose gels and the DNA extracted using a gel extraction kit (Qiaquick, Qiagen). Sequencing of PCR amplified fragments was done by the UBC Nucleic Acid Protein Service (NAPS) using Applied Biosystem (ABI) amplitaq Dyedeoxy terminator sequencing techniques and ABI prism 337 automatic sequencers. The fragments were sequenced from each end using the same primers as used in PCR. The computer program Se-Al 2.1 was used to assist in manual alignment and comparison of the sequences.

3. The experiments: Specific Methods and Results

3.1 Background investigations

Several preliminary investigations were required. I needed to eliminate the possibility that mutations in *comA* prevent binding and uptake of DNA by affecting the expression of competence genes. It was also essential to show that my experimental methodology could be expected to work. The gene responsible for resistance to novobiocin was used to demonstrate that *H. influenzae* could be transformed by PCR products, and that the mutagenesis created point mutations as expected. Finally, the colony competence assay needed to be investigated to determine whether or not the age of colonies affects competence.

3.1.1 *comA* does not alter expression of other competence genes

A protein that regulates the expression of competence genes is unlikely to be a part of the uptake machinery. Working with Ph.D candidate Andrew Cameron, I created strains to confirm that *comA*, as well as *comC,* and *comE,* are not regulatory.

rec-2 is expressed during the development of competence [85]. Using RR871, which has a duplicate *rec-2* gene fused with *lacZ*, expression levels of *rec*-2 can be inferred by measuring β-galactosidase activity. In a cell that is unable to become competent due to a regulatory defect we expect that *rec-2* will not be expressed, and β-galactosidase activity will be low.

We created cells with both mutant *comA* and *rec2::lacZ*-cm^R by transforming KW20 simultaneously with RR871 and JG48 (*comA::*tn10-kan^R) DNA, then selecting for both kanamycin and chloramphenicol resistance. The transformation was repeated using DNA from JG7 (*comC::*tn10-Kan^R), and JG1 (*comE::*tn10-Kan^R) instead of JG48 to create isolates mutant at both *comC* and *comE*. Two transformant colonies of each were cultured and a 15 ml sample transferred to MIV as per the competence-inducing protocol. The remainder of the sBHI cultures continued growing overnight. β-galactosidase activity was assessed from samples taken in log growth, stationary phase (overnight cultures), and MIV-treated cells following the methods described in [30]. RR871 was included as a positive control in all conditions. RR885 (*cya*$^-$, *rec-2::lacZ*-cm^R) was included to indicate the reduced activity expected in regulation mutants in all conditions, and MIV-induced KW20 was included as a control.

The average β-galactosidase activity for each strain and condition is shown in Figure 5. As only two replicates were done, statistical analysis is not possible. In all conditions, the double mutants of *com* gene knockouts with *rec-2::lacZ* showed β-galactosidase activity roughly equal to the activity in RR871. The *cya*$^-$, *rec-2::lacZ* control had severely reduced β-galactosidase activity, as expected. These results indicate that the expression of *rec-2* is not influenced by the competence gene mutations.

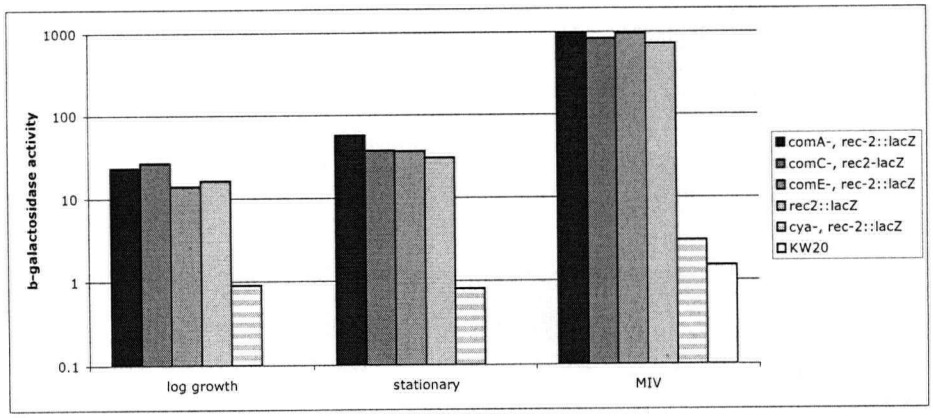


Figure 5: β-galactosidase activity in *com/rec*-2 double mutants.

If the loss of DNA binding and uptake in the competence mutants is not due to a regulatory effect, it follows that the proteins are either a part of the binding and uptake mechanism, or assist in assembly of the components of the mechanism.

3.1.2 *H. influenzae* can be transformed by PCR products

As the uptake of DNA by *H. influenzae* is sequence specific, competent cells should readily take up PCR products. However, this had not yet been demonstrated. My experiment protocol involved transforming cells with mutated PCR fragments, so it was necessary to do this.

To test the ability of *H. influenzae* to be transformed by PCR fragments, the novobiocin resistance conferring *gyrB* allele (*gyrB*nov^R) from MAP7 was used. Primers mbgyrB1 (5'-CTCTTTGGTGCCCTTTCAGTCAT) and mbgyrB2 (5'-CCTTTTTTTATCGTTTTCCTTTCC) were designed to flank the entire *gyrB* sequence. The resulting PCR fragment was 2564 bp in length. Four ten-fold dilutions of the PCR product were used to transform competent KW20 cells to novobiocin resistance, at concentrations 16 ng/ml to 0.016 ng/ml. Each transformation was performed in duplicate.

Transformation frequencies were expected to decrease proportionately to the tenfold dilution steps over the lower concentrations, possibly reaching saturation in the higher concentrations. Figure 6 shows the transformation frequency to novobiocin resistance for each dilution of the MAP7 *gyrB* DNA, as well as the no DNA control and the

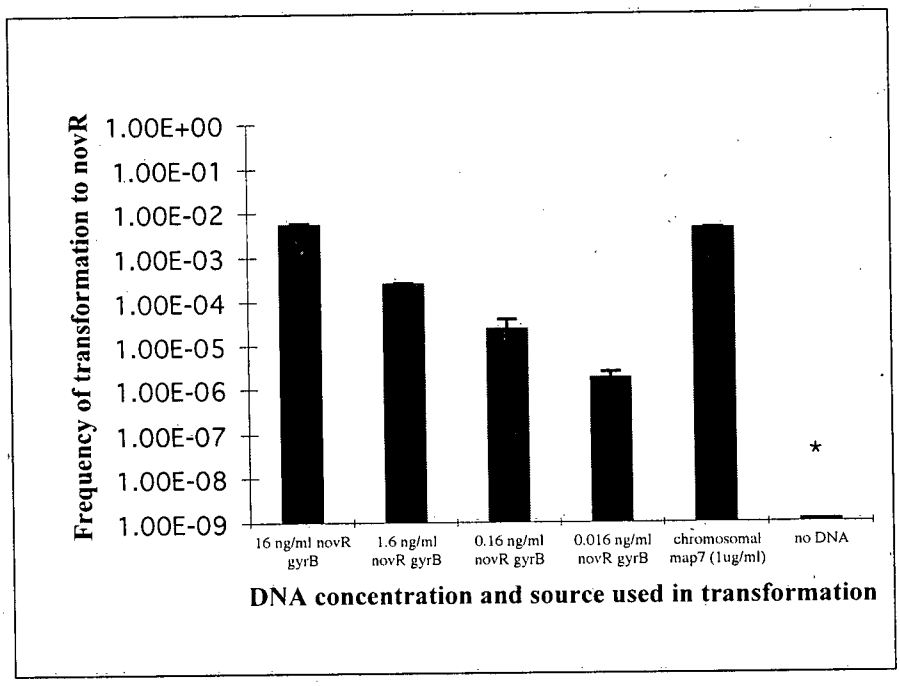


Figure 6: Transformation using PCR fragments. Transformation Frequencies of MIV competent KW20, using PCR amplified *gyrB* from MAP7 (nov^R), compared to Transformation Frequencies using genomic MAP7 DNA. Error bars indicate range of duplicates. * indicates that colonies were seen only in one duplicate.

chromosomal MAP7 DNA control. The maximal transformation frequency of the PCR products was roughly equal to that of MAP7 chromosomal DNA. Each ten-fold dilution resulted in a proportional decrease in transformation frequency. This indicates that 16 ng/ml was not a saturating concentration of DNA.

3.1.3 The mutation protocol creates point mutations in *gyrB*

To test the mutational ability of Clontech's mutagenic PCR kit, I used the kit to amplify *gyrB*nov^S under mutagenic conditions. Point mutations in a highly conserved ATPase region of *gyrB* interfere with novobiocin binding to DNA gyrase, conferring resistance to this antibiotic. Transformation of KW20 with *gyrB* amplified under the kit's mutagenic conditions would indicate that the kit is functioning, if there was an increase in novobiocin resistant colonies. It would then follow that the kit will produce point mutations in *comA*.

The Clontech mutagenic PCR kit is expected to create point mutations during amplification. Specific point mutations in *gyrB* have been mapped in novobiocin resistant *E. coli, S. pneumoniae*, and *S. aureus* [86-89]. The number of nov^R sites in GyrB depends on the species, but by comparing sequences in the ATPase region, it can be estimated that any one of up to four amino acid substitutions could lead to a nov^R phenotype in *H. influenzae*. Predictions can then be made of the expected frequency of mutation to nov^R under mutagenic PCR conditions, and are shown in the prediction bars for each mutation rate in Figure 7. The probability of any base being substituted by a particular base is one-third the mutation rate per Kb, divided by 1000. The probability of

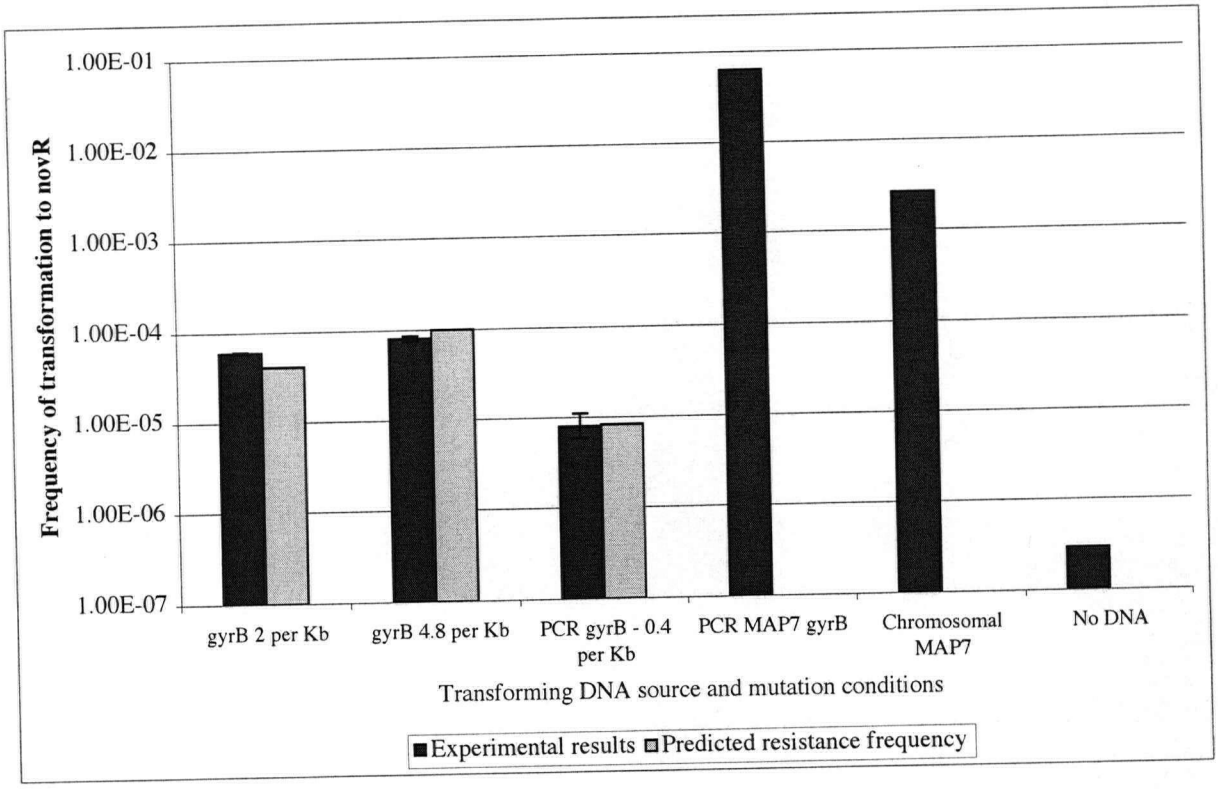


Figure 7: Mutation of *gyrB*(nov^S) to *gyrB*(nov^R). Frequencies of transformation of MIV competent KW20 to novobiocin resistance, using wild-type *gyrB* amplified under mutagenic conditions. Error bars indicate duplicate ranges.

a cell taking up and being transformed by a nov^R fragment is then the probability of the fragment containing this mutation multiplied by the probability of transformation based on transformation by the amplified *gyrB*nov^R fragment. This calculation makes the assumptions that there is only one site at which a mutation will prevent novobiocin binding to the protein, that only one specific substitution at that location will suffice, and that the probability of mutation to any given base is equally likely.

Using primers from the previous experiment, I used the kit to amplify the *gyrB*nov^S allele from KW20 DNA under conditions predicted to produce 2.0 and 4.8 substitutions per Kb. To confirm that the conditions were appropriate for mutation, I followed the protocol for the kit's controls. As well, I amplified *gyrB*nov^S in normal PCR conditions, expected to have a mutation rate of 0.4 substitutions per Kb according to the supplier. Competent KW20 cells were transformed with the *gyrB* mutational PCR products and with the products of the normal PCR amplification, each at transformation concentrations of approximately 16 ng/ml. The positive controls were transformation with the *gyrB*nov^R PCR products from the previous experiment and with MAP7 chromosomal DNA. Each transformation was performed in duplicate.

The results of the transformations by mutagenized *gyrB* are shown in Figure 7. In all mutation rate categories the frequency of transformation to nov^R closely matched the predictions (although statistical analysis is not reasonable due to the low number of replicates). The transformation frequency of the culture with chromosomal MAP7 DNA was 2.6×10^{-3}, within normal range of MIV transformation frequency of KW20 (approximately 1×10^{-2} for nov^R). The transformation frequency using normal fidelity

PCR amplified *gyrB*nov^R was higher ($6.2x10^{-2}$). This is a higher transformation frequency than obtained earlier with similar quantities of DNA. However, the concentration was possibly greater than the estimated 16 ng/ml. The kit's control amplifications indicated that in the control tubes, the conditions were appropriate for mutation (not shown).

It is evident that the mutation protocol is producing point mutations in *gyrB*, and at a rate within realistic expectations, and it can be expected that this technique will produce point mutations in *comA*.

3.1.4 Colony competence decreases in RR804 after 36 hours

To maximize the sensitivity of the colony competence assay used to identify possible reduced transformation frequency mutants, I examined the effect of colony age on competence.

Cultures of KW20 and the hypercompetent RR804 were plated at dilutions appropriate to yield individual colonies. Two colonies were sampled from each at intervals over three days, and their competence determined by the colony competence assay.

The changes in transformation frequencies of the colonies with age are shown in Figure 8. Maximal transformation frequencies of each were comparable to published data [47]

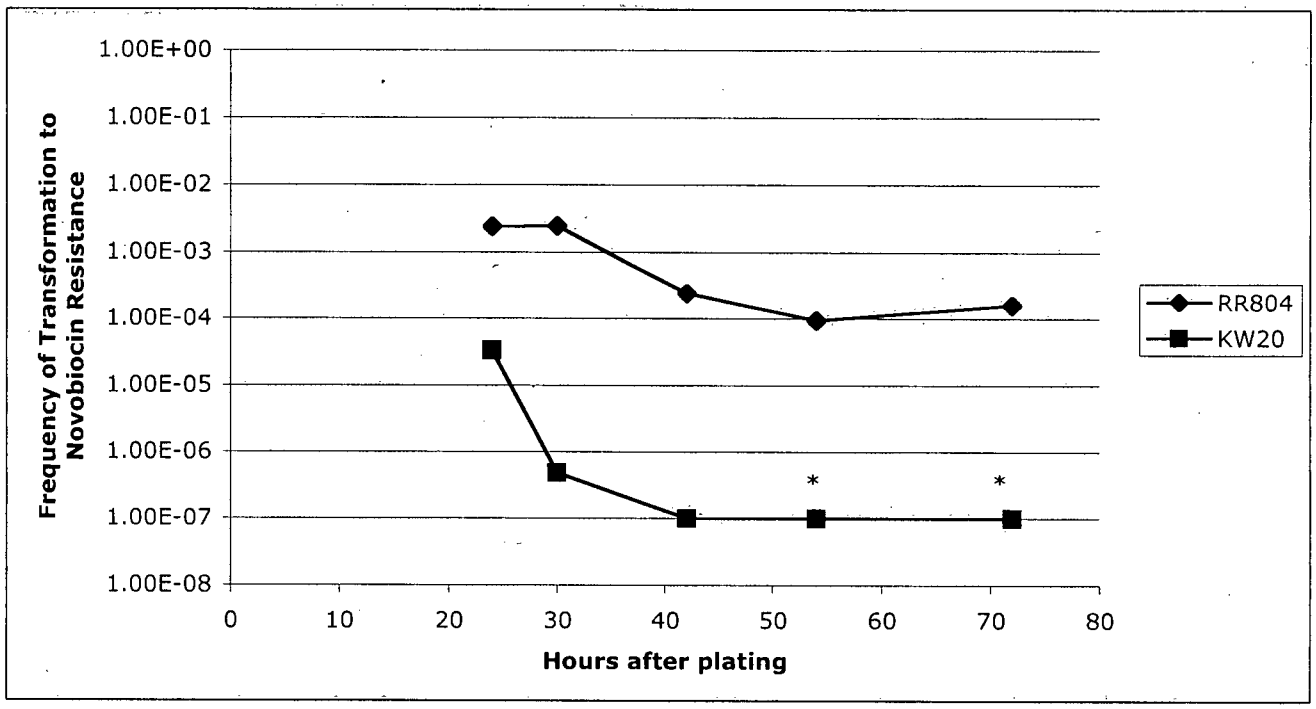


Figure 8: Colony competence timecourse. * indicates value less than or equal to that shown because no colonies were seen.

in 24 hour-old colonies. At 35 hours, the transformation frequency of KW20 had decreased, and transformation was barely detectable at 42 hours. After this point, no transformation was detected. RR804 maintained maximal transformation frequency at 30 hours, then decreased. This indicates that to ensure maximum potential transformation frequencies, RR804 colonies should be used in colony competence assays within 30 hours of plating.

3.2 Mutation of *comA* and co-transformation

Once it was confirmed that PCR fragments of amplified *gyrB* could transform *H. influenzae*, the mutagenesis experiments with *comA* could proceed. Primers comAF (5'-AGCAAGCAAGATAAATAAAGCGATA) and comAR (5'-GAGACCTATGATTTGGCTGTTAAGT) were designed to lie outside of *comA* as shown in Figure 9, so the 950 bp PCR fragment includes the entire gene. *comA* was amplified using these primers, under conditions expected to produce a mutation rate of 4.8 point mutations per Kb, as well as under normal PCR conditions. As *comA* is 798 bp in length, the mutagenic condition should produce an average of 3.8 substitutions within the gene.

PCR products were checked using gel electrophoresis. Each reaction resulted in a single band that migrated with the 1Kb band of the ladder. The kit's mutagenesis controls were included, and indicated that mutagenic conditions were appropriate in the control tubes. These results predict that *comA* was successfully amplified and mutated in the mutagenesis reaction.

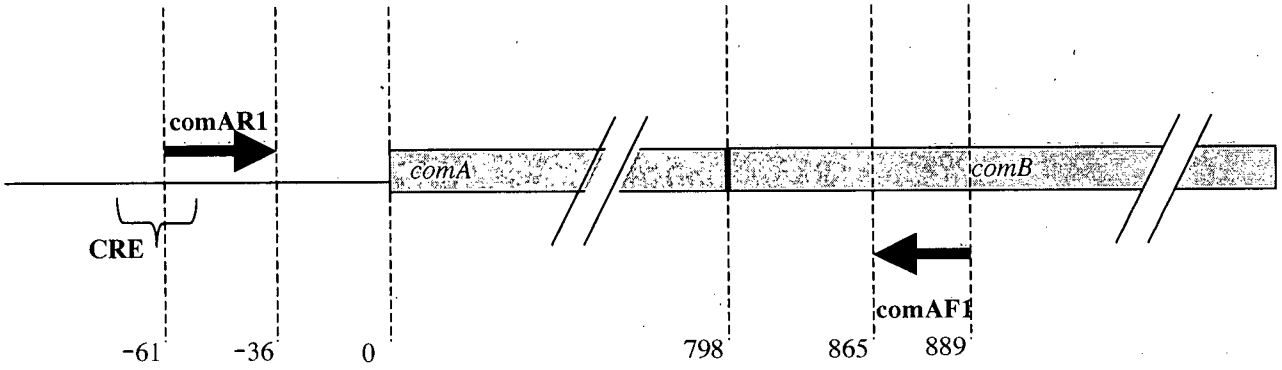


Figure 9: *comA* amplification. Location of primers in relation to *comA*, *comB*, and the CRE site upstream of the operon.

The mutated *comA* fragment does not contain a selectable marker, and so it was impossible to detect which cells had taken up *comA* in a normal transformation procedure. Expected transformation frequencies are approximately $1x10^{-2}$, and so only roughly one in a hundred colonies would contain a mutant *comA*. To increase the likelihood that transformed colonies contained a mutated *comA* I used a co-transformation procedure. Cells were transformed with the selectable *gyrB*nov^R PCR amplified fragment at a limiting concentration (1:23) compared with the excess concentration of the mutated *comA* fragment. As cells take up multiple fragments, a cell transformed to novobiocin resistance is then likely to also have a mutant *comA* fragment.

RR804 was transformed in 500 µl aliquots with 23 ng *comA4.8* (*comA* amplified at a mutation rate of 4.8 point mutations per Kb) and 1ng *gyrB*nov^R, then plated on novobiocin plates. As determining transformation frequency was not important to this stage, DNaseI was not used; production of sufficient transformed colonies was my goal. As controls, 20 ng *gyrB*nov^R and 500 µg of chromosomal MAP7 DNA were each used to transform RR804 cells to novobiocin resistance, and a no DNA control was used to confirm that spontaneous mutation to novobiocin resistance was rare. The PCR reactions and cotransformations were performed multiple times to provide colonies for use on several days, as the following screen was time-consuming.

3.3 Screening for reduction in transformation frequency

To identify mutations in *comA* that caused a phenotype of altered recognition of the uptake signal sequence, I searched for mutations that reduced transformation. A series of

two screens were used. Both screens are modifications of the colony competence assay, outlined in the Materials and Methods chapter. The first screen identified possible reduced-competence mutants, the second and more sensitive screen was used to eliminate erroneously identified cell lines.

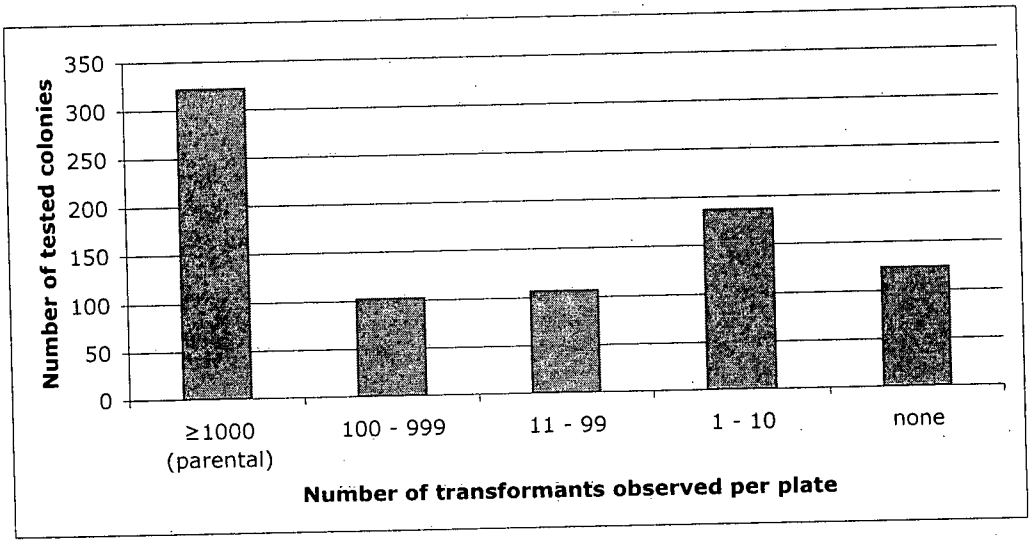
3.3.1 Initial Screen

In the first screen, colonies from the *comA*4.8/*gyrB*nov^R transformation were tested 24-30 hours after transformation and plating. Single colonies were picked from the plate and then touched to a numbered location on a catalogue plate before being suspended in sBHI containing MAP7 DNA. The assay proceeded as described in the Materials and Methods chapter. Fifty microlitres of the transformation suspension was plated on kanamycin plates. As a positive control, four colonies of RR804 transformed with *gyrB*nov^R were assayed.

847 colonies of *comA*4.8*/*gyrB*(nov^R) were screened, shown in Figure 10 a. The majority of colonies had transformation frequencies indistinguishable from RR804 colonies, resulting in more than 1000 transformant colonies per plate. Colonies that showed no transformation (zero colonies growing on the kanamycin plates) were selected for the second screen.

125 colonies had zero transformants per plate, but only 81 of these were determined to be possible low transformation frequency mutants, because there were several instances where more than ten colonies sampled in a row produced no transformants in the assay.

A



B

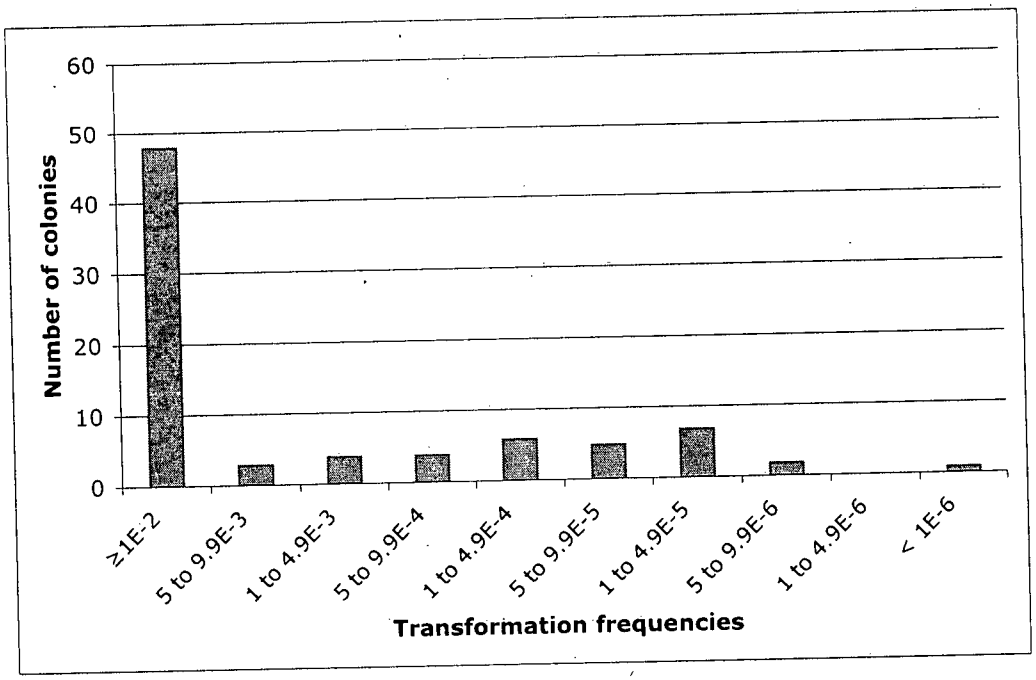


Figure 10: Identification of mutants with reduced transformation frequency. (**A**) First screen: the distribution of transformability after mutagenesis (871 colonies tested).(**B**) Second screen: re-test of 81 colonies from 'none' category of first screen. The parental strain had a TF of 1.3E-2.

As this appeared suspicious, thirteen of these colonies from two series were subjected to the assay a second time, and showed normal (above 1000 colonies) transformation rates. It was assumed that these were representative of other similar series, and I concluded that experimental error or variation in plates was responsible. The source colonies of these series were not included in the subsequent screen.

3.3.2 Second Screen

The 81 reduced transformation frequency colonies were streaked on fresh plates from the catalogue plate colonies. After 24-36 hours, four colonies of each isolate were picked and separately resuspended in 1ml of sBHI with MAP7 DNA. The reduced volume, compared with the initial screen, increases the density of cells and thus the sensitivity of the test. The suspensions were plated on kanamycin plates, and diluted and plated on plain plates, allowing calculation of transformation frequency. Each time this screen was performed, four colonies of RR804 transformed with *gyrB*nov^R were subjected to the same test, as positive controls.

The results of this secondary screen are shown in Figure 10 b. Of these 81 cell lines, 48 were found to have a transformation frequency comparable to RR804 transformed with *gyrB*nov^R. Twenty-five colonies had transformation frequencies at least 10 fold below this baseline transformation frequency. The remainder had transformation frequencies between the expected transformation frequency for RR804 and the ten-fold reduction.

Each screened colony is expected to contain a mutated *comA*. Roughly three percent of all screened colonies showed a decrease in transformation frequency of at least tenfold. This is likely an underestimate, as it is possible that some colonies that were not subjected to the secondary screen also had this low transformation frequency.

3.4 Sequencing of *comA* from mutants

The *comA* gene of each of the 25 lines with the lowest transformation frequencies (at least 10-fold below RR804 levels) was sequenced to determine the locations of the mutations resulting in this phenotype. *comA* alleles from RR804 and KW20 were also sequenced. The sequencing primers were those used in mutagenesis, so the sequenced region contained the entire mutagenized segment. The PCR products were gel-purified prior to sequencing.

Sequencing was done in both directions, to create overlap in the middle region. Mutant sequences were compared to the wild type and RR804, as well as to the published sequence for KW20 *comA* available from the TIGR website. Ambiguities were investigated using the electropherograms supplied by the sequencing center.

Although an average of 3.8 mutations was expected, none were found in the 25 mutants. This is not due to poor sequence data. The sequences of the *comA* fragment provided clear unambiguous sequences, with approximately 300 bp overlap. In regions where the signal was insufficient or there was overlap of peaks the complementary reverse strand confirmed that there were no differences from wild type. The only apparent difference

from the published sequences was immediately after comAR1, but this was consistent across all isolates including KW20. This was likely a misread error, based on the consistency of the error. The fluorescent signal in the sequencing process was high, and may have been strong enough in the initial bases of the sequence to saturate the sequencing apparatus.

3.5 Investigation of Mutants

The reduced transformation frequencies of the 25 isolates must be due to mutations elsewhere in the genome. The location of the mutations was unknown, and mapping their locations would be too time-consuming and difficult for this study. However, two experiments were done to further characterize the isolates.

3.5.1 *MurE* is unchanged

The parental strain was hypercompetent due to a point mutation in *murE (murE749)*. A reversion of this point mutation would cause cells to have normal KW20 competence, which would appear in the colony competence assay as a dramatic reduction in transformation frequency. This can be checked because the hypercompetence point mutation created a new *Mnl I* site.

To check for retention of the *murE749* allele, a 587 bp region surrounding the *murE* point mutation was amplified from colonies of each mutant, KW20, and RR804 using Primer 1

and Primer 2 [47]. The fragments were then digested with *Mnl1* for 3 hours and separated in a 4% agarose gel.

The PCR amplification produced the expected 587 bp band, as well as a faint second band at 340 bp which was ignored (gel not shown). Figure 11 shows representatives of the restriction fragment gels. The restriction fragments of KW20 and RR807 migrated as expected. KW20 has a single *Mnl I* cut site, and was cut into two fragments, one 219 bp, the other 368. RR807 has a second *Mnl I* cut site, resulting in three fragments of 219, 229, and 139 bp. All twenty-five mutants showed bands that migrate similarly to the RR804 pattern, with a thick double band near 220 bp and one near 140 bp.

These results indicate that the hypercompetent RR804 mutation is still present in all twenty-five mutant isolates. The mutations causing their reduction in transformation frequency must be located elsewhere.

3.5.2 MIV with limiting DNA

In a normal tranformation assay, cells are given an excess of DNA, but using limited DNA for transformation allows greater sensitivity. The presence of excess DNA may mask subtle differences in competence, allowing a cell with reduced uptake ability to still be transformed at a similar rate to the wild type. If DNA is present at low concentrations, binding DNA is a less frequent event, and defects in transformation will be more obvious.

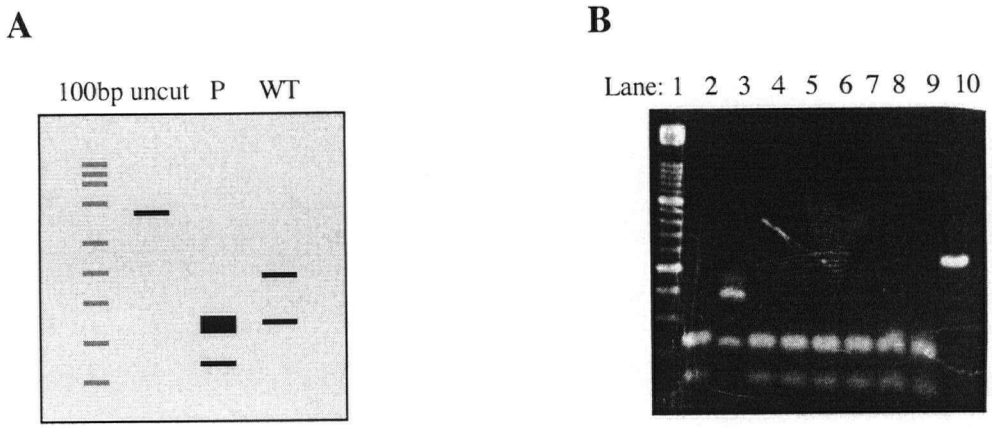


Figure 11: Digestion of *murE* fragment with *MNLI*. **(A)** Diagram of expected band sizes. WT fragment sizes should be 219 and 368 bp. Hypercompetent mutation of *murE* (P) should be cut to 219, 229 and 139 bp. **(B)** Sample of six mutants compared to WT and P. Lane 1: 100 bp ladder, 2: RR804, 3: KW20, 4 through 9: mutants, 10: uncut KW20 fragment.

Each of the 25 reduced-competence isolates was transformed with limiting DNA (10 ng/1ml culture) after 100 min incubation in MIV. RR804 was also transformed to show the expected maximal transformation frequency under these conditions. MIV competent cells of the wild type KW20 strain were also transformed with excess DNA (1 μg/ml culture) to ensure that the conditions were sufficient to induce maximal competence. This was conducted in triplicate, and all but eight isolates had transformation frequencies similar to the wild type. The eight isolates with low transformation frequencies were assayed a second time in duplicate.

The results of the second test are shown in Figure 12. KW20 with excess DNA had a transformation frequency of $1.5x10^{-2}$, indicating that the conditions were appropriate to fully induce competence. With limited DNA, RR804 and KW20 had transformation frequencies of $2.86x10^{-4}$ and $2.78x10^{-4}$, respectively. This approximately 100-fold decrease in transformation frequency corresponds with the decrease in DNA available. Four of the eight isolates re-tested had transformation frequencies ten-fold lower than their parental strain; colonies 21, 22, 23, and 18 had transformation frequencies $3.43x10^{-5}$, $3.21x10^{-5}$, $4.56x10^{-5}$ and $1.62x10^{-5}$ respectively.

These results indicate that the four isolates with a consistently low frequency of transformation possess mutations that affect their ability to efficiently bind DNA.

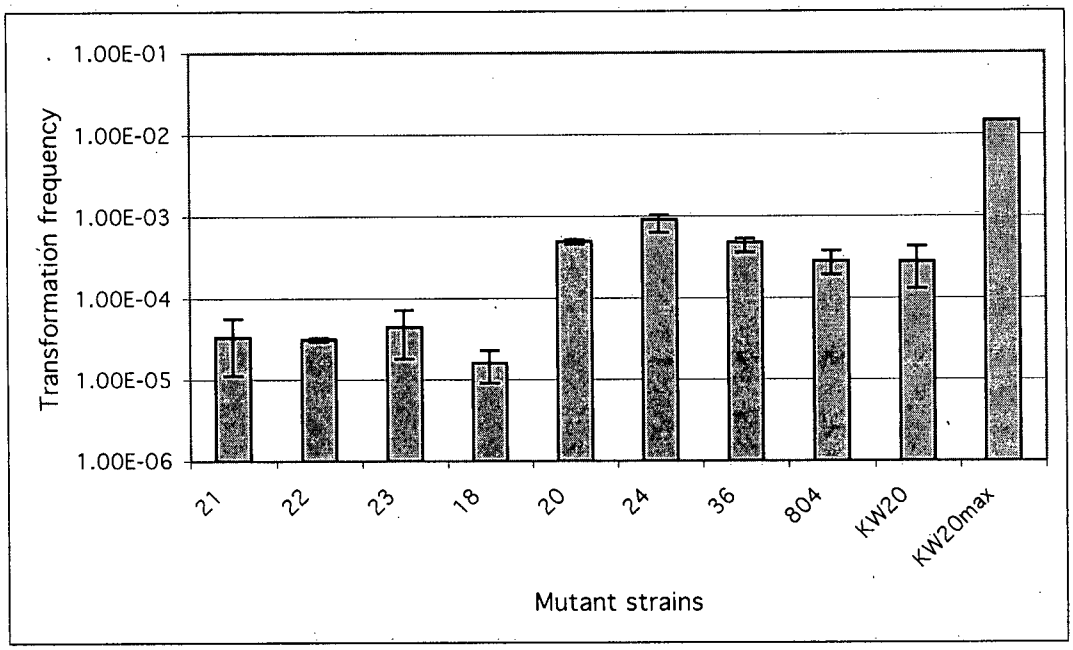


Figure 12: Transformation frequencies of selected mutants under MIV-induced, limiting DNA conditions. Mutant strains are identified by their original catalogue number. KW20max represents transformation of MIV-competent KW20 cells with non-limited concentrations of DNA. Error bars show ranges of duplicates.

3.6 Is transformation mutagenic?

It is possible that transformation itself is mutagenic. None of the isolates had mutations in *comA*, yet they had mutations that affected transformation frequency. Finding twenty-five mutants out of 847 screened indicates a high rate of mutation, and yet the only apparent treatment was transformation by *comA*4.8 and *gyrB*nov^R.

To investigate the mutagenic effects of transformation, I transformed RR804 with three different combinations of PCR products and screened for reduced transformation frequency mutants using the same initial screen in my *comA* experiment. The four transformation treatment groups were: (A) 1 ng *gyrB*nov^R, 25ng *comA*4.8; (B) 1 ng *gyrB*nov^R, 25 ng *comA*; (C) 26 ng *gyrB*nov^R; and (D) no DNA. 44 colonies of A, C, and D, and 45 colonies of B were screened.

In the no DNA treatment, all tested colonies had normal transformation frequencies. No plates had obviously reduced numbers of colonies. The three transformed treatment groups each had similar numbers of colonies with reduced transformation frequencies. This is shown in Figure 13.

The transformed treatment groups all had rates of reduced transformation frequency similar to those seen in the first screen of mutagenized *comA* transformed colonies. This indicates that the uptake of DNA or transformation is causing mutations in the cells.

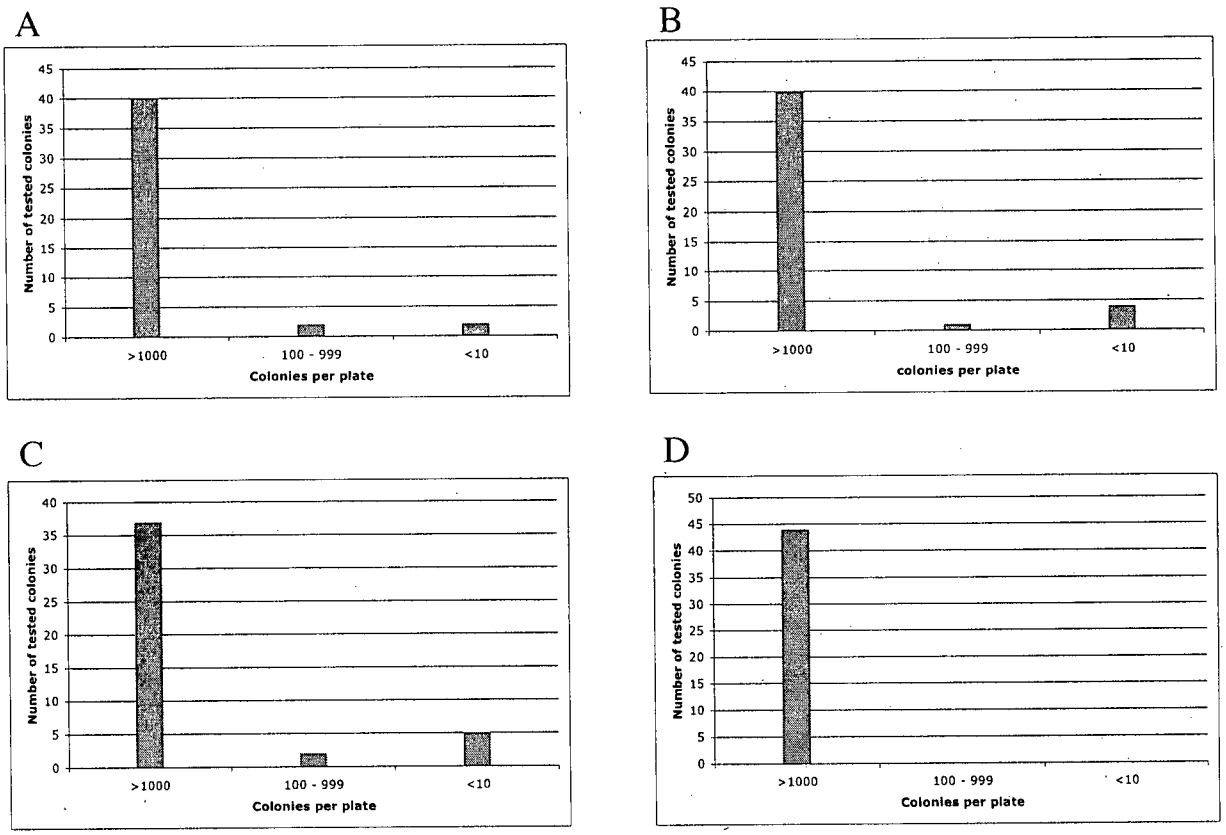


Figure 13 : Mutagenic effects of transformation. (A) 1 ng *gyrB*nov^R, 25 ng *comA*4.8; (B) 1 ng *gyrB*nov^R, 25 ng *comA*; (C) 26 ng *gyrB*novR; and (D) no DNA.

4. Discussion

4.1 Background investigations

The background investigations provided basic information as to the role of the *com* operon genes in DNA binding, demonstrated that the protocol I planned to use in mutagenesis should result in mutation of *comA*, and optimized sensitivity of the labour-intensive mutant screen.

Mutations in *comA*, *comC* and *comE* did not cause loss of expression of *rec-2*, another gene expressed in competence development. This showed that the mutant *com* operon genes do not cause loss of DNA binding by interfering with expression of the DNA binding protein. Had *comA* played a regulatory role, it would not have been suitable for my research question.

The mutagenesis protocol used in my experiments required that competent *H. influenzae* cells could be transformed by PCR amplified fragments. PCR reagents, at the concentrations used in the amplification reaction, should not affect competence nor viability when added to a culture. The uptake mechanism is sequence-based, and so PCR amplified fragments should be taken up efficiently as long as they contain at least one USS. The ability of *H. influenzae* to be transformed by PCR fragments was experimentally confirmed, as *gyrB* amplified from novobiocin resistant cells efficiently transformed sensitive cells to resistance. This amplified fragment had a higher

transformation frequency than genomic DNA. This is not unexpected, as only a small fraction of genomic fragments contain the novobiocin resistance allele.

The critical part of the initial experiments was the confirmation that the mutagenesis protocol could produce mutations at the desired rate, and that *H. influenzae* would express a mutant phenotype in accordance with these mutations after transformation. In the mutation of *gyrB* experiment, the frequency of transformation to resistance was approximately ten fold higher when low-mutation conditions were used to amplify *gyrB*, compared with normal PCR conditions. This corresponds with the ten-fold differences in the expected mutation rates: 4.8 mutations per Kb, compared with 0.4 mutations per Kb. Without sequencing *gyrB* from novobiocin resistant transformants, it is not possible to know if the mutation rates were exactly as the mutagenesis kit claimed, but it is evident that mutation was occurring, and at approximately the desired rate.

Colony age was shown to effect competence in both KW20 and RR804. In order to ensure maximal levels of competence, and in turn sensitivity of the mutant screen, transformed RR804 colonies were screened for competence in later tests between 24 and 30 hours of plating. This meant that for each day of screening for mutants, new plates of colonies were needed.

4.2 Identification of mutants in transformation

Point mutations in a gene may offer more information about a protein than can be gained from insertion/deletion mutations. The relevance of specific amino acids to the function of the protein can be detected, or the function itself can be altered, instead of simply the phenotype associated with a complete lack of function. To attempt an analysis of the function of *comA*, mutagenic PCR was used. EMS would also have produced point mutations, but they would occur at random throughout the entire genome. Using PCR allowed me to target *comA*, with primers that flanked this gene.

I used PCR conditions designed to produce a mutation rate of 4.8 per Kb, a rate higher than suggested for protein function studies. Vartanian *et al.* [71] suggest that mutation rates producing on average 1.5 mutations per gene are useful for studying wild type gene function. Because the genetic code is redundant, some point mutations will not affect the amino acid sequence. Introducing just one base substitution per copy of *comA* would be inefficient. This is a method that can be used when attempting to identify individual amino acids crucial to protein function, but in my study I was interested in interrupting or altering function, intending to change the ability of the cell to recognize the USS. Alteration in more than one amino acid should not hinder this goal. Using a rate of 4.8 mutations per Kb should have introduced an average of 3.7 substitutions in each *comA* fragment. Had I isolated mutants with more than one substitution in *comA* , it would have been uncertain which of the mutations caused the reduction in transformation frequency. Further investigation to identify which mutation was responsible would give information on functional regions within the protein.

Prediction of the frequency at which mutations of a given type should be seen was problematic. Little is known of the structure of ComA, and nothing of the relationship between structure and function. Predicting the effect of an amino acid substitution within the protein is therefore nearly impossible. However, it should be possible to estimate the frequency of stop codons generated by the mutations. Three of sixty-four mutations should result in a stop codon. Knowing the length of *comA* (795 bp), the expected frequency of mutations (4.8/Kb, resulting in a predicted 3.816 mutations per gene), and the fact that 3/64 of these should produce a stop codon, the frequency at which a stop codon should arise can be calculated to be 0.1788. This does not directly translate into a 'knockout' phenotype, as it may be that the protein can retain partial or complete function in the absence of a portion of its C-terminus. Even so, given this frequency and the number of colonies screened in my study, it would be reasonable to expect that at least one knockout mutation should be observed. However, out of 847 colonies screened, all had detectable transformation frequencies either initially or on the second screen.

When *comA* was sequenced from each mutant line, it was seen that no mutations of any type had occurred in this gene. Clearly, mutations in *comA* were not responsible for the reduced transformation frequency phenotype. This was unanticipated, as spontaneous mutations should be rare, and so the only differences between the mutants and RR804 should be in *gyrB* and *comA*.

The consequence of these results was that the original purpose of these experiments had to be abandoned. Mutations in *comA* could not be used to determine the role of ComA in USS binding, nor to test hypotheses of the evolution of USS recognition.

4.3 Investigation of mutants

Though the mutants identified in the colony competence screens did not have mutations in *comA*, they are still mutants defective in the ability to be transformed. The mutations responsible for the reduced transformation frequency phenotypes may be located within genes already identified as competence genes, within genes with other functions that have an effect on competence, or in genes not yet identified as being competence genes. It was beyond the scope of this experiment to map the location of the mutations responsible for the reduction in transformation frequency, however some possible locations were tested, and the nature of the defect in competence was explored.

4.3.1 Did the *murE* mutation revert?

The parental strain was hypercompetent due to a point mutation within *murE* [47]. It was not surprising that none of the 25 tested reduced competence isolates were reversions to the wild type version of *murE*, as this would be unlikely.

Though the mutations responsible for the decreases in transformation frequency were not due to a back-mutation of the *murE* point mutation, they could still be mutations that

interfere with the unknown *murE* pathway to competence. These could be located in *murE*, elsewhere in the murein pathway, or in some other regulatory gene.

4.3.2 MIV competence of mutants

Four mutant isolates had transformation frequencies lower than RR804 under MIV competence inducing conditions with limiting DNA concentrations. These four probably have mutations in genes directly involved in some step of the competence pathway. This does not necessarily mean they are involved in the mechanism of DNA uptake; they could be regulatory or part of the recombination apparatus.

The seventeen isolates with MIV/limiting DNA transformation frequencies comparable to RR804 may have been identified due to experimental error. The consistency of the phenotype in the two colony competence assay tests weakens this argument. It is unlikely that a colony was accidentally identified twice as having reduced transformation frequency. These seventeen isolates may have mutations in genes that affect the pathway by which the *murE* mutation turns on competence. Four of the eight re-tested isolates returned to transformation frequency levels comparable to RR804. These may have been unstable mutations, or it may be that the phenotypes are sensitive to subtle environmental cues.

4.4 What went wrong?

4.4.1 Why were there no mutations in *comA*?

It is surprising that not one mutation in *comA* was isolated. Mutation of *gyrB* produced point mutations that conferred novobiocin resistance to cells transformed by this gene. This indicates that the mutation and transformation protocol should work. We know that *comA* possesses a USS, and that the tested cells that should have been transformed with *comA* were competent. Had they not been, they would not have been transformed to novobiocin resistance in the cotransformation step. As *comA* was in 25-fold excess of *gyrB*, it is unlikely that the mutants only bound and took up *gyrB*. The number of fragments taken up by competent cells has been estimated to be between four and forty, dependent on size of fragment [20]. Point mutations in *comA* are unlikely to be lethal, because the knockout is fully viable.

Clues to the reason why the mutant *comA* did not transform the cells might lie in the differences between the DNA fragments themselves. Perhaps the *comA* fragment forms a secondary structure that inhibits uptake, or perhaps the effect of two USSs in *gyrB* outweighs the abundance of the *comA* fragment in the cotransformation.

The size difference between the fragments may also be responsible; *gyrB* is three times the length of *comA*. Once within the cell, the single strand of DNA is possibly protected

from degradation by a protein such as SSB, but it does slowly degrade. When creating knockouts, or attempting to make other non-homologous insertions, a rule of thumb that is used is that there should be five hundred bp of homologous flanking DNA to either side of the insertion. This suggests that much less than five hundred bases are degraded, leaving well over half the fragment in the case of *comA*. Barany *et al.* [22] suggest, however, that up to 1.5 Kb of a fragment is degraded prior to recombination. It is then a possibility that the *comA* fragment was either completely degraded before recombination could occur, or was too short to recombine.

This reason for the lack of transformation by mutated *comA* could be tested using *gyrB*. We know the general region of this gene in which point mutations cause novobiocin resistance. Primers could be designed to flank this region, resulting in an amplified fragment approximately 1 Kb in length, then this DNA could be used to transform competent cells. A decrease in transformation frequency (relative to transformation frequency of the entire novR *gyrB* gene), or a total loss of transformation, would indicate that a 1 Kb fragment is not long enough to be reliably integrated.

4.4.2 Why were there any mutants?

The process that created the mutations isolated in this study is also an unknown. The appearance of decreased transformation frequency mutants in all transformation groups (section 3.6) suggests that transformation itself may cause mutations. The lack of such mutants in the no-DNA treatment group indicates that this phenotype is not simply

present at low levels in the RR804 population, nor is it the expression of competence genes that causes the mutations. It is possible that the process of recombination or the presence of single stranded DNA in the cell increases the mutation rate.

There are other possible explanations. Recent research suggests that mutation rates increase in bacteria that have been exposed to low levels of some antibiotics, reviewed in [90]. It is possible that a cell that has the MAP7 originating *gyrB* mutation perceives 2.5 μg/ml novobiocin as being a comparably low level. This concentration is sufficient to prevent growth of non-resistant cells, but perhaps resistant cells are slightly susceptible rather than insensitive to novobiocin. This may be sufficient to increase mutation rates.

Another cause of increased mutation rates could have been introduced alongside the mutated *comA* DNA. The *gyrB* used to transform the cells was PCR amplified, and as already discussed, PCR has a low but not inconsequential mutation rate. Mutations in the C-terminal half of the *gyrB* gene have been implicated in alterations of mutation rates. One mutation within *gyrB* is known to increase mutation rates, while another decreases them, by altering the extent of the negative supercoiling [91]. However, this would take several steps to achieve; a more direct explanation is that the alterations in negative supercoiling affect other functions in the cell, such as transcription and recombination. In this way, it would act much as a *topA* knockout does, interfering with transformation due to changes in DNA coiling [92].

4.5 Future directions

Though the study failed to yield any *comA* mutants, this does not mean the mutants that were isolated are useless. They affect the ability of cells to be transformed, and may offer new knowledge of the system. The mutations may be located in genes already described, which would offer the same opportunity to study those genes as point mutations in *comA* would have offered. They may equally be located in genes as yet only suspected of involvement in competence, or novel genes. The mutations may also not be within a gene, but may be located in a regulatory sequence instead. However, mapping an unknown mutation is time-consuming. It may be better to first fully characterize the mutant phenotypes.

Various techniques can be used to identify the process affected by the mutation. If the mutation affects regulation of competence gene expression, real-time PCR can be used to quantify transcription of a competence gene. Deficiencies in binding and uptake may be detected using DNAse treatments after incubation with labeled DNA; cells that can bind and take up DNA will have label associated with them after DNAse treatment, while cells that cannot bind DNA will not.

Though the mutations might have occurred in any gene involved in competence, there is a possibility that one or more of the mutant strains has a mutation in the gene responsible for USS binding. Such a mutation would be expected to alter the cell's ability to selectively take up *H. influenzae* DNA. The increase in the ability of *H. influenzae* to take up foreign DNA could be detected by a competition assay [93]. DNA containing the

USS competes with other USS containing DNA for binding and uptake. Diluting *H. influenzae* DNA with *E.coli* in a wild type cell does not alter the transformation frequency of the cell by a marker on the conspecific DNA. If recognition of the USS is altered, *E.coli* DNA would then be in competition with *H. influenzae* DNA, and transformation frequency of that marker would decrease.

A competition assay was attempted with the four mutant strains (data not shown). However, the transformation frequencies of the wild type control with varying concentrations of the two types of DNA were inconsistent. No conclusions could be drawn. It is possible that the DNA I was using was contaminated, or that concentrations were not accurate.

The questions of where the mutations came from, and why *comA* was not transformed, are as interesting as the mutations themselves. A better-designed experiment to investigate the source of the mutations may be worthwhile, as it may be that PCR generated mutations in *gyrB* could affect mutation rates. To circumvent this, RR804 transformed with chromosomal MAP7 DNA should be used. To investigate the lack of *comA* mutations, several possible experiments may be done. Uptake studies using labeled *comA* would demonstrate whether *comA* is in fact being taken up into the cell. If it is being taken up, the question is then 'why doesn't it recombine?' It may be that the fragment was too short, in which case a longer fragment could be amplified.

4.6 Conclusions

The attempt at identifying the gene responsible for USS binding by mutation of *comA* was unsuccessful, as the procedure used to mutate the gene failed. This means that there is no new information on the role of ComA in competence, and no progress towards understanding the recognition of the USS.

There are now four new competence mutants available for study, each with a reduction in transformation frequency phenotype. At this time, it is not considered a high priority to map or further characterize these mutants, but it is possible that they will be useful in the future. These new strains may have independent mutations in already identified competence genes, or may have mutations in genes not known at this point to be involved in competence.

References

1. Kroll, J.S., et al., *Natural genetic exchange between Haemophilus and Neisseria: intergeneric transfer of chromosomal genes between major human pathogens.* Proc Natl Acad Sci U S A, 1998. **95**(21): p. 12381-5.
2. Kroll, J.S., E.R. Moxon, and B.M. Loynds, *Natural genetic transfer of a putative virulence-enhancing mutation to Haemophilus influenzae type a.* J Infect Dis, 1994. **169**(3): p. 676-9.
3. Goodgal, S.H. and M.A. Mitchell, *Sequence and uptake specificity of cloned sonicated fragments of Haemophilus influenzae DNA.* J Bacteriol, 1990. **172**(10): p. 5924-8.
4. Goodman, S.D. and J.J. Scocca, *Identification and arrangement of the DNA sequence recognized in specific transformation of Neisseria gonorrhoeae.* Proc Natl Acad Sci U S A, 1988. **85**(18): p. 6982-6.
5. Moxon, E.R. and R. Wilson, *The role of Haemophilus influenzae in the pathogenesis of pneumonia.* Rev Infect Dis, 1991. **13 Suppl 6**: p. S518-27.
6. Quagliarello, V. and W.M. Scheld, *Bacterial meningitis: pathogenesis, pathophysiology, and progress.* N Engl J Med, 1992. **327**(12): p. 864-72.
7. Mwangi, I., et al., *Acute bacterial meningitis in children admitted to a rural Kenyan hospital: increasing antibiotic resistance and outcome.* Pediatr Infect Dis J, 2002. **21**(11): p. 1042-8.
8. Murphy, T., *Haemophilus infections*, in *Harrisons principles of internal medicine*, Braunwald F, et al., Editors. 2001, McGraw Hill: New York. p. 939 - 942.
9. Fleischmann, R.D., et al., *Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.* Science, 1995. **269**(5223): p. 496-512.
10. Lorenz, M.G. and W. Wackernagel, *Bacterial gene transfer by natural genetic transformation in the environment.* Microbiol Rev, 1994. **58**(3): p. 563-602.
11. Solomon, J.M. and A.D. Grossman, *Who's competent and when: regulation of natural genetic competence in bacteria.* Trends Genet, 1996. **12**(4): p. 150-5.
12. Dubnau, D., *DNA uptake in bacteria.* Annu Rev Microbiol, 1999. **53**: p. 217-44.
13. Chen, I. and E.C. Gotschlich, *ComE, a competence protein from Neisseria gonorrhoeae with DNA-binding activity.* J Bacteriol, 2001. **183**(10): p. 3160-8.
14. Smeets, L.C., et al., *The dprA gene is required for natural transformation of Helicobacter pylori.* FEMS Immunol Med Microbiol, 2000. **27**(2): p. 99-102.
15. Hofreuter, D., S. Odenbreit, and R. Haas, *Natural transformation competence in Helicobacter pylori is mediated by the basic components of a type IV secretion system.* Mol Microbiol, 2001. **41**(2): p. 379-91.
16. Palmen, R. and K.J. Hellingwerf, *Uptake and processing of DNA by Acinetobacter calcoaceticus--a review.* Gene, 1997. **192**(1): p. 179-90.
17. Smith, H.O., M.L. Gwinn, and S.L. Salzberg, *DNA uptake signal sequences in naturally transformable bacteria.* Res Microbiol, 1999. **150**(9-10): p. 603-16.
18. Smith, H.O., et al., *Frequency and distribution of DNA uptake signal sequences in the Haemophilus influenzae Rd genome.* Science, 1995. **269**(5223): p. 538-40.
19. Aas, F.E., C. Lovold, and M. Koomey, *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, 2002. **46**(5): p. 1441-50.
20. Barouki, R. and H.O. Smith, *Initial steps in Haemophilus influenzae transformation. Donor DNA binding in the com10 mutant.* J Biol Chem, 1986. **261**(19): p. 8617-23.
21. Goodgal, S.H., *DNA uptake in Haemophilus transformation.* Annu Rev Genet, 1982. **16**: p. 169-92.
22. Barany, F., M.E. Kahn, and H.O. Smith, *Directional transport and integration of donor DNA in Haemophilus influenzae transformation.* Proc Natl Acad Sci U S A, 1983. **80**(23): p. 7274-8.
23. Chaussee, M.S. and S.A. Hill, *Formation of single-stranded DNA during DNA transformation of Neisseria gonorrhoeae.* J Bacteriol, 1998. **180**(19): p. 5117-22.

24. Pifer, M.L. and H.O. Smith, *Processing of donor DNA during Haemophilus influenzae transformation: analysis using a model plasmid system.* Proc Natl Acad Sci U S A, 1985. **82**(11): p. 3731-5.
25. Herriott, R.M., E.M. Meyer, and M. Vogt, *Defined nongrowth media for stage II development of competence in Haemophilus influenzae.* J Bacteriol, 1970. **101**(2): p. 517-24.
26. Williams, P.M., L.A. Bannister, and R.J. Redfield, *The Haemophilus influenzae sxy-1 mutation is in a newly identified gene essential for competence.* J Bacteriol, 1994. **176**(22): p. 6789-94.
27. Redfield, R.J., M.R. Schrag, and A.M. Dean, *The evolution of bacterial transformation: sex with poor relations.* Genetics, 1997. **146**(1): p. 27-38.
28. Boling, M.E., J.K. Setlow, and D.P. Allison, *Bacteriophage of Haemophilus influenzae. I. Differences between infection by whole phage, extracted phage DNA and prophage DNA extracted from lysogenic cells.* J Mol Biol, 1972. **63**(3): p. 335-48.
29. Setlow, J.K., et al., *Relationship between prophage induction and transformation in Haemophilus influenzae.* J Bacteriol, 1973. **115**(1): p. 153-61.
30. MacFadyen, L.P., et al., *Competence development by Haemophilus influenzae is regulated by the availability of nucleic acid precursors.* Mol Microbiol, 2001. **40**(3): p. 700-7.
31. Neuhard, J. and R.A. Kellin, *Biosynthesis and conversions of pyrimidines*, in *Escherichia coli and Salmonella: cellular and molecular biology*, F. Neidhardt, Editor. 1996, ASM Press: Washington. p. chapter 35.
32. Zalkin, H. and P. Nygaard, *Biosynthesis of purine nucleotides*, in *Escherichia coli and Salmonella: cellular and molecular biology*, F. Neidhardt, Editor. 1996, ASM Press: Washington. p. chapter 34.
33. Macfadyen, L.P., et al., *Regulation of competence development and sugar utilization in Haemophilus influenzae Rd by a phosphoenolpyruvate:fructose phosphotransferase system.* Mol Microbiol, 1996. **21**(5): p. 941-52.
34. Dorocicz, I.R., P.M. Williams, and R.J. Redfield, *The Haemophilus influenzae adenylate cyclase gene: cloning, sequence, and essential role in competence.* J Bacteriol, 1993. **175**(22): p. 7142-9.
35. Macfadyen, L.P., *Regulation of competence development in Haemophilus influenzae.* J Theor Biol, 2000. **207**(3): p. 349-59.
36. Redfield, R.J., *Do bacteria have sex?* Nat Rev Genet, 2001. **2**(8): p. 634-9.
37. Michod, R.E., M.F. Wojciechowski, and M.A. Hoelzer, *DNA repair and the evolution of transformation in the bacterium Bacillus subtilis.* Genetics, 1988. **118**(1): p. 31-9.
38. Redfield, R.J., *Evolution of natural transformation: testing the DNA repair hypothesis in Bacillus subtilis and Haemophilus influenzae.* Genetics, 1993. **133**(4): p. 755-61.
39. Redfield, R.J., et al., *A novel CRP-dependent regulon controls expression of competence genes in Haemophilus influenzae.* J Mol Biol, 2005. **347**(4): p. 735-47.
40. Aas, F.E., et al., *Competence for natural transformation in Neisseria gonorrhoeae: components of DNA binding and uptake linked to type IV pilus expression.* Mol Microbiol, 2002. **46**(3): p. 749-60.
41. Corpet, F., J. Gouzy, and D. Kahn, *The ProDom database of protein domain families.* Nucleic Acids Res, 1998. **26**(1): p. 323-6.
42. Servant, F., et al., *ProDom: automated clustering of homologous domains.* Brief Bioinform, 2002. **3**(3): p. 246-51.
43. Nielsen, H., et al., *Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites.* Protein Eng, 1997. **10**(1): p. 1-6.
44. Wang, Y., et al., *Natural transformation and DNA uptake signal sequences in Actinobacillus actinomycetemcomitans.* J Bacteriol, 2002. **184**(13): p. 3442-9.
45. Wang, Y., et al., *Type IV pilus gene homologs pilABCD are required for natural transformation in Actinobacillus actinomycetemcomitans.* Gene, 2003. **312**: p. 249-55.
46. Chandler, M.S., *The gene encoding cAMP receptor protein is required for competence development in Haemophilus influenzae Rd.* Proc Natl Acad Sci U S A, 1992. **89**(5): p. 1626-30.
47. Ma, C. and R.J. Redfield, *Point mutations in a peptidoglycan biosynthesis gene cause competence induction in Haemophilus influenzae.* J Bacteriol, 2000. **182**(12): p. 3323-30.
48. Tomb, J.F., et al., *Transposon mutagenesis, characterization, and cloning of transformation genes of Haemophilus influenzae Rd.* J Bacteriol, 1989. **171**(7): p. 3796-802.

49. Tomb, J.F., H. el-Hajj, and H.O. Smith, *Nucleotide sequence of a cluster of genes involved in the transformation of Haemophilus influenzae Rd.* Gene, 1991. **104**(1): p. 1-10.
50. Zulty, J.J. and G.J. Barcak, *Identification of a DNA transformation gene required for com101A+ expression and supertransformer phenotype in Haemophilus influenzae.* Proc Natl Acad Sci U S A, 1995. **92**(8): p. 3616-20.
51. Karudapuram, S., X. Zhao, and G.J. Barcak, *DNA sequence and characterization of Haemophilus influenzae dprA+, a gene required for chromosomal but not plasmid DNA transformation.* J Bacteriol, 1995. **177**(11): p. 3235-40.
52. Drake, S.L. and M. Koomey, *The product of the pilQ gene is essential for the biogenesis of type IV pili in Neisseria gonorrhoeae.* Mol Microbiol, 1995. **18**(5): p. 975-86.
53. Rost, B. and C. Sander, *Prediction of protein secondary structure at better than 70% accuracy.* J Mol Biol, 1993. **232**(2): p. 584-99.
54. Larson, T.G. and S.H. Goodgal, *Sequence and transcriptional regulation of com101A, a locus required for genetic transformation in Haemophilus influenzae.* J Bacteriol, 1991. **173**(15): p. 4683-91.
55. Larson, T.G. and S.H. Goodgal, *Donor DNA processing is blocked by a mutation in the com101A locus of Haemophilus influenzae.* J Bacteriol, 1992. **174**(10): p. 3392-4.
56. Brewster, J.M. and E.A. Morgan, *Tn9 and IS1 inserts in a ribosomal ribonucleic acid operon of Escherichia coli are incompletely polar.* J Bacteriol, 1981. **148**(3): p. 897-903.
57. Siehnel, R.J. and E.A. Morgan, *Efficient read-through of Tn9 and IS1 by RNA polymerase molecules that initiate at rRNA promoters.* J Bacteriol, 1983. **153**(2): p. 672-84.
58. Franklin, N.D. and C. CYanofsky, *The N protein of lambda: evidence bearing on transcription termination, polarity and the alternation of E.coli RNA polymerase*, in *RNA polymerase*, R. Losick and M. Chamberlin, Editors. 1976, Cold Spring Harbour Laboratory: Cold Spring Harbour, NY. p. 693 - 706.
59. Molnar, S., 2004. Identification and characterization of DNA receptor candidates in *Haemophilus influenzae*. Thesis for Master of Science. The University of British Columbia, Vancouver.
60. VanWagoner, T.M., et al., *Characterization of three new competence-regulated operons in Haemophilus influenzae.* J Bacteriol, 2004. **186**(19): p. 6409-21.
61. Inamine, G.S. and D. Dubnau, *ComEA, a Bacillus subtilis integral membrane protein required for genetic transformation, is needed for both DNA binding and transport.* J Bacteriol, 1995. **177**(11): p. 3045-51.
62. Ma, C., Unpublished.
63. Dougherty, B.A. and H.O. Smith, *Identification of Haemophilus influenzae Rd transformation genes using cassette mutagenesis.* Microbiology, 1999. **145**(Pt 2): p. 401-9.
64. Finkel, S.E. and R. Kolter, *DNA as a nutrient: novel role for bacterial competence gene homologs.* J Bacteriol, 2001. **183**(21): p. 6288-93.
65. Keohavong, P. and W.G. Thilly, *Fidelity of DNA polymerases in DNA amplification.* Proc Natl Acad Sci U S A, 1989. **86**(23): p. 9253-7.
66. Andre, P., et al., *Fidelity and mutational spectrum of Pfu DNA polymerase on a human mitochondrial DNA sequence.* Genome Res, 1997. **7**(8): p. 843-52.
67. Dietrich, J., et al., *PCR performance of the highly thermostable proof-reading B-type DNA polymerase from Pyrococcus abyssi.* FEMS Microbiol Lett, 2002. **217**(1): p. 89-94.
68. Cadwell, R.C. and G.F. Joyce, *Randomization of genes by PCR mutagenesis.* PCR Methods Appl, 1992. **2**(1): p. 28-33.
69. Fromant, M., S. Blanquet, and P. Plateau, *Direct random mutagenesis of gene-sized DNA fragments using polymerase chain reaction.* Anal Biochem, 1995. **224**(1): p. 347-53.
70. Clontech, *Diversify PCR Random Mutagenesis Kit User Manual.* 2001.
71. Vartanian, J.P., M. Henry, and S. Wain-Hobson, *Hypermutagenic PCR involving all four transitions and a sizeable proportion of transversions.* Nucleic Acids Res, 1996. **24**(14): p. 2627-31.
72. Melnikov, A. and P.J. Youngman, *Random mutagenesis by recombinational capture of PCR products in Bacillus subtilis and Acinetobacter calcoaceticus.* Nucleic Acids Res, 1999. **27**(4): p. 1056-62.
73. Karlin, S., J. Mrazek, and A.M. Campbell, *Frequent oligonucleotides and peptides of the Haemophilus influenzae genome.* Nucleic Acids Res, 1996. **24**(21): p. 4263-72.

74. Bakkali, M., et al., *Evolutionary stability of DNA uptake signal sequences in the Pasteurellaceae.* Proc Natl Acad Sci U S A, 2004. **101**(13): p. 4513-8.
75. Sourice, S., et al., *Identification of the Chi site of Haemophilus influenzae as several sequences related to the Escherichia coli Chi site.* Mol Microbiol, 1998. **27**(5): p. 1021-9.
76. Wilcox, K.W. and H.O. Smith, *Isolation and characterization of mutants of Haemophilus influenzae deficient in an adenosine 5'-triphosphate-dependent deoxyribonuclease activity.* J Bacteriol, 1975. **122**(2): p. 443-53.
77. Alexander, H. and G. Leidy, *Determination of inherited traits of H.influenzae by deoxyribonucleic acid fractions isolated from type-specific cells.* J Exp Med, 1951. **93**: p. 345-349.
78. Goodgal, S.H., *Procedures for Haemophilus influenzae transformation.* Methods Enzymol, 1968. **12**: p. 867-876.
79. Catlin, B.W., J.d. Bendler, and S.H. Goodgal, *The type b capsulation locus of Haemophilus influenzae: map location and size.* Journal of General Microbiology, 1972. **70**(3): p. 411-22.
80. Gwinn, M.L., et al., *Role of the two-component signal transduction and the phosphoenolpyruvate: carbohydrate phosphotransferase systems in competence development of Haemophilus influenzae Rd.* J Bacteriol, 1996. **178**(21): p. 6366-8.
81. Redfield, R.J., *sxy-1, a Haemophilus influenzae mutation causing greatly enhanced spontaneous competence.* J Bacteriol, 1991. **173**(18): p. 5612-8.
82. Poje, G. and R.J. Redfield, *Transformation of Haemophilus influenzae.* Methods Mol Med, 2003. **71**: p. 57-70.
83. Wishart, D.S., P. Stothard, and G.H. Van Domselaar, *PepTool and GeneTool: platform-independent tools for biological sequence analysis.* Methods Mol Biol, 2000. **132**: p. 93-113.
84. Engels, B., *Amplify 2.1.* 1992, Department of Genetics, University of Wisconsin: Madison, WI.
85. Gwinn, M.L., et al., *In vitro Tn7 mutagenesis of Haemophilus influenzae Rd and characterization of the role of atpA in transformation.* J Bacteriol, 1997. **179**(23): p. 7315-20.
86. Munoz, R., M. Bustamante, and A.G. de la Campa, *Ser-127-to-Leu substitution in the DNA gyrase B subunit of Streptococcus pneumoniae is implicated in novobiocin resistance.* J Bacteriol, 1995. **177**(14): p. 4166-70.
87. Stieger, M., et al., *GyrB mutations in Staphylococcus aureus strains resistant to cyclothialidine, coumermycin, and novobiocin.* Antimicrob Agents Chemother, 1996. **40**(4): p. 1060-2.
88. Fournier, B. and D.C. Hooper, *Mutations in topoisomerase IV and DNA gyrase of Staphylococcus aureus: novel pleiotropic effects on quinolone and coumarin activity.* Antimicrob Agents Chemother, 1998. **42**(1): p. 121-8.
89. Contreras, A. and A. Maxwell, *gyrB mutations which confer coumarin resistance also affect DNA supercoiling and ATP hydrolysis by Escherichia coli DNA gyrase.* Mol Microbiol, 1992. **6**(12): p. 1617-24.
90. Blazquez, J., A. Oliver, and J.M. Gomez-Gomez, *Mutation and evolution of antibiotic resistance: antibiotics as promoters of antibiotic resistance?* Curr Drug Targets, 2002. **3**(4): p. 345-9.
91. Setlow, J.K., D. Haines, and E. Cabrera-Juarez, *Gyrase mutants affect mutation in a localized region of Haemophilus influenzae.* Mutat Res, 2001. **478**(1-2): p. 83-8.
92. Chandler, M.S. and R.A. Smith, *Characterization of the Haemophilus influenzae topA locus: DNA topoisomerase I is required for genetic competence.* Gene, 1996. **169**(1): p. 25-31.
93. Deich, R.A. and H.O. Smith, *Homologous and heterologous DNA uptake in Haemophilus transformation*, in *Transformation*. 1978. p. 377-384.