INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

Preparatory Studies Towards the Identification of Genes Regulated by the Ner-like Protein (Nlp) of *Escherichia coli*

Submitted by: Stephanie L. Matheson
9648502

Submitted as partial fulfillment of a Master of Science degree
Department of Microbiology and Immunology
McGill University
Montreal, Quebec, Canada
November, 1998

Abstract

Bacterial gene regulation often occurs at the transcriptional level. Several different DNA binding motifs have been identified in transcriptional regulatory proteins from both prokaryotes and eukaryotes. Comprising a unique family of DNA binding proteins, with a conserved DNA binding domain, are the negative early repressor (Ner) of temperate coliphages Mu and D108, the Ner-like protein (Nlp) of *E.coli*, and the HIV-1 TATA element modulatory factor (TMF) of humans.

During the lytic cycle of growth, the Ner proteins of the transposable coliphages Mu and D108 negatively regulate the transcription of the *c* repressor by binding to their respective operator sites, thereby blocking transcription from the repressor/promoter P_c. In doing so , Ner also effectively turns off its own expression, plus the expression of the other early genes from P_e required for the lytic cycle. Transcription from P_e is also positively regulated by binding of the host protein IHF which acts to overcome H-NS mediated repression.

Nlp shares approximately 60% amino acid sequence homology with Mu and D108 Ner, having higher homology to both proteins than the proteins share themselves. Nlp cannot bind to Ner binding sites (of either Mu or D108) *in vitro*, nor does it confer pseudoimmunity to either phage *in vivo*.

Originally identified by its ability to complement a *crp*1* mutant of *E.coli* (MK2001) with respect to maltose fermentation, the *nlp* gene was found to be located at 69.3 minutes on the *E.coli* chromosome and to encode for a 91 amino acid polypeptide. In addition, Northern blotting and primer extension analysis have indicated that the *nlp* gene is monocistronic, with its start site 29 base pairs upstream of its ATG start codon. Through *nlp* insertion mutations with *lac* and *luxAB* reporter genes, it was also determined that this gene is expressed but not essential for cell viability.

Since Nlp function is relatively unknown, a *lacZ* transcriptional fusion library is one approach to identify genes whose expression is controlled by Nlp. As a preliminary study of the use of such methodology, a library was constructed without the cloned gene of interest (*ie. nlp*). To do this, the *nlp* gene was interrupted using *luxAB* and a tetracycline resistance gene in the Δ *lac* strain *E. coli* 40. This strain (LF20300) was subsequently transformed with an expression vector possessing the *ara* P_{BAD} promoter and a chloramphenicol resistance marker (pBAD18Cm). P_{BAD} vectors allow for activation of the P_{BAD} promoter by the addition of arabinose to the growth media, and repression of its expression by the addition of glucose to the growth media. The plasmid-containing strain was then lysogenized with a Mu d1 phage containing the *lacZYA* reporter gene and an ampicillin resistance marker. The resulting lysogens were selected on TCMG plates containing 0.2% glucose, tetracycline, chloramphenicol, and ampicillin. These lysogens were then mastered onto TCMG plates containing either 0.2% glucose or 0.2% arabinose plus X-gal and the appropriate antibiotics. Differences in blue colour were noted for eight different clones on glucose in comparison to their growth on arabinose. Total genomic DNA was isolated from each arabinose-responsive clone, digested with several different restriction endonucleases, and subjected to Southern hybridization (using labeled *lacZ* as a probe). Putative restriction maps were determined for those clones with single Mu d1 inserts, and used to identify candidate genes from the complete restriction map of *E. coli*. By examining the acquisition of artifacts and arabinose inducible genes, one may ascertain the effectiveness of the future construction of a similar library to identify Nlp-responsive genes.

RESUME

La régulation de l'expression des gènes bactériens se produit souvent au moment de la transcription. Différents sites de reconnaissance de l'ADN, impliqués dans le phénomène de régulation transcriptionnelle des protéines, ont pu être identifiés tant chez les procaryotes que les eucaryotes. Par exemple, le répresseur négatif (Ner) des coliphages Mu et D108, le Nlp *d'E.coli* et le facteur de modulation du VIH-1 TATA humain font partie d'une famille unique de protéines de reconnaissance de l'ADN.

Au cours du cycle de croissance lytique, les protéines Ner des coliphages transposables Mu et D108 modulent négativement la transcription du répresseur *c* en se liant a leurs sites actifs. Il inhibe ainsi la transcription du répresseur promoteur Pc. Finalement, Ner supprime sa propre expression et celle des gènes natifs de Pe impliqués dans le cycle lytique. La transcription de Pe est positivement modulée par la formation d'une liaison avec la protéine hôte HIF dont le rôle est d'empêcher la répression médiée par H-NS.

Nlp présente 60 % d'homologie de séquence d'acides aminés avec Mu et D108 Ner. Soulignons que l'homologie de séquence est plus marquée entre Nlp et les deux protéines Mu ou D108 Ner que ces deux protéines sont homologues entre elles. *In vitro*, Nlp ne peut pas se lier aux sites de reconnaissance de Ner (que ce soit Mu ou D108), de même qu'il ne confère, *in vivo*, aucune pseudo-immunité a ces phages.

Le Nlp a été mis en evidence à partir de sa capacité à completer l'action du *crp*1* mutant *d'E.coli* (MK2001) permettant la fermentation du maltose. Le Nlp a été localisé sur le chromosome *d'E.coli* a 69,3 minutes et code pour un polypeptide comportant 91 acides aminés. De plus, la technique du Northern blot et celle d'analyse de l'extension des initiateurs ont démontré que le gène *nlp* est monocistronique avec un site d'initiation situé a 29 paires de bases avant le codon ATG de départ. En utilisant des mutations d'insertion sur le gène *nlp* par introduction de gènes de type *lac* et *luxAB*, il a été démontré que ce gène est exprimé mais n'est cependant pas essentiel à la survie de la cellule.

Etant donné que la fonction du Nlp est encore malconnue, l'une des approches pour l'identification des gènes, dont l'expression est controlé par Nlp, est la création d'une librairie de fusion transcriptionnelle lacZ. Le but de cette préétude étant la mise au point d'une méthodologie, cette librairie a été construite sans utilisation du gène clôné qui nous interesse (*nlp*). A cet effet, lèxpression du gène nlp a été inhibé par insertion de luxAB et du gène de résistance à la tétracycline dans *E.coli 40* (souche Δ-*lac*). Cette souche (LF20300) a été par la suite transformée avec un vecteur d'expression contenant le promoteur ara P_{BAD} et un marqueur resistant chloramphenicol (pBAD18Cm). Les vecteurs de type P_{BAD} permettent l'activation du promoteur par addition d'arabinose dans le milieu de croissance et l'inhibition de son expression par addition de glucose dans ce même milieu. La souche avec plasmide a été ensuite lysogènisé avec un phage Mu dl comportant le gène de lecture *lacZYA* et un marqueur de résistance ampicilline.

Les lysogènes résultants ont été sélectionnés en utilisant des plaques contenant 2% de glucose, de la tétracycline, du chloroamphénicol et de l'ampiciline. Ces lysogènes ont ensuite été cultivées sur des plaques TCMG contenant du glucose (0.2%) ou de l'arabinonse (0.2%), du X-gal ainsi que l'antibiotique approprié. Des différences de couleur entre les deux milieux de cultures (glucose et arabinose) ont été remarquées pour 8 clones. La totalité de l'ADN génomique a été isolée pour chaque clone coloré en présence d'arabinose, digérée par plusieurs enzymes de restriction et soumise à l'analyse d'hybridation de Southern en utilisant le lacZ radioactif comme sonde. Des cartes de restriction ont été déterminées pour ces clones en faisant une mono-insertion du gène Mu dl. Elles ont été aussi utilisées pour identifier les gènes à l'étude, dans la carte de resctriction d'*E. Coli*. L'examen de la formation de gènes induits et non-induits par la présence d'arabinose démontre l'efficacité éventuelle de la construction d'une librairie similaire pour analyser les gènes induits ou réprimés par la protéine Nlp.

Acknowledgements

I wish to thank everyone who assisted me during the past two years, especially my supervisor Dr. Michael DuBow, my lab mates (past and present), Caroline Diorio (our generous technician), and Dr. Gina MacIntyre (a former postdoctoral fellow). Without all of you, I would still be floundering in a sea of confusion. I extend special thanks to Christine Yew, Clare Jennings and Sally Bockler (my partner in crime, and in THALLIUM! Productions) for making my transition to Montreal an extremely pleasant experience. If it hadn't been for you, I would have returned home long ago. Many thanks also to my dear friends Wendy Osborn and Hong Hua for the occasional pep talks, brief visits, and encouraging words. Thank you to Dr. Jon Beckwith at Harvard University, for graciously providing me with the pBAD series of vectors. I also wish to thank the coffee bean growers of Columbia, and the manufacturers of Excedrin™. Lastly, and most importantly, I thank my beloved family: Mum, Dad, Brent, Christen, and Duchess (good dog!), you have been of tremendous spiritual support. Thank you with all of my heart.

"An unlesson'd girl, unschool'd, unpractised,
Happy in this, she is not yet so old
But she may learn."

William Shakespeare
(Act III, Scene II, The Merchant of Venice)

Table of Contents

List of Figures and Tables

Chapter I

Chapter II

Chapter III

List of Abbreviations

amp :	ampicillin
cam :	chloramphenicol
dNTP :	deoxynucleotide triphosphate
kan :	kanamycin
LB :	Luria Bertani
^{R} :	resistant
tet :	tetracycline
Tn :	transposon
X-gal :	5-bromo-4-chloro-3-indoyl-β-D-galactoside

Chapter I: General Introduction

CHAPTER I: General Introduction

Methods for the Study of Gene Expression

Both prokaryotic and eukaryotic gene expression is controlled at several different levels such as transcription, translation, RNA processing, mRNA degradation, and protein activity (Neidhart *et.al.*, 1996). The regulation of gene expression is crucial for survival, and mirrors a myriad of changes in the pathway from genome to mRNA translation. Alterations in gene expression often occur due to external factors, and specialized cell types may result from differential expression induced by such signals (Alberts *et.al.*, 1994).

One of the most commonly addressed themes in the field of modern genetics is the relationship between genotype and phenotype (Chee *et.al.*, 1996). Genetic variation, amongst populations and individuals, most likely relates a functional diversity. Therefore, it is essential to understand not only DNA sequences within genomes, but also the expression of encoded genes. Various methods have been devised whereby this relationship may be elucidated. For example, one can measure a fusion gene product (in either a transcriptional or translational fusion library), mRNA levels (since mRNA is the most recognized intermediate in gene expression), or quantities of expressed protein (Silhavy and Beckwith, 1985).

Within the prokaryotic realm of existence, there is a relatively economic use of genomic sequence. Unlike eukaryotes, bacteria maintain a limited amount of "junk DNA". Since the life cycle of a microbe requires more genetic adaptability than that of its multicellular counterpart, the correlation between DNA sequence and gene expression is

intriguing. Now that entire sequence for *Escherichia coli* and other bacteria have been published, perhaps this mystery will gradually unravel. Moreover, the marriage of knowledge and technology will simplify the task.

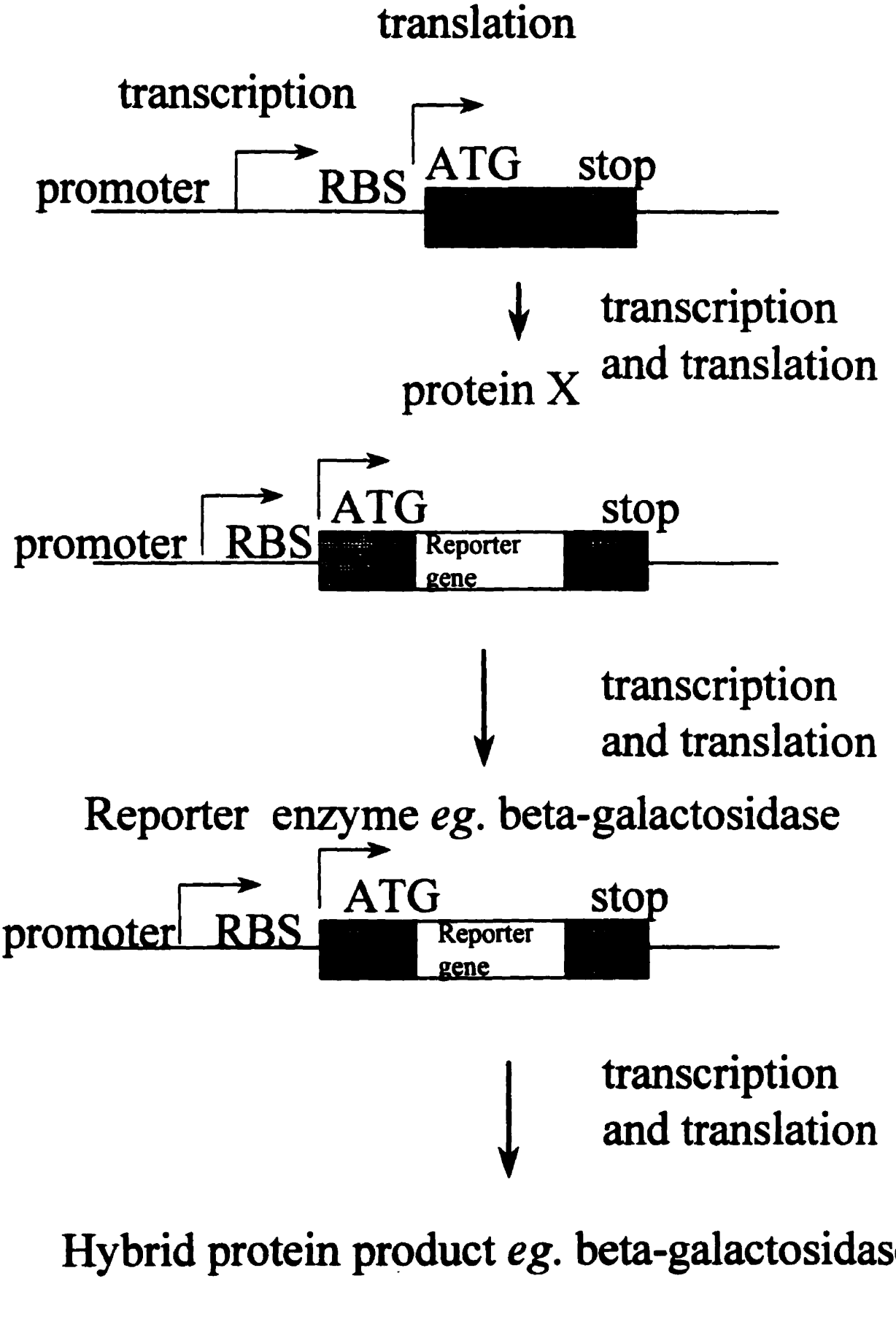
For the purpose of this review, the focus will be on the study of gene expression in prokaryotes (and its relation to function) using two different methods: transcriptional and fusion libraries, two-dimensional protein electrophoresis gels, and oligonucleotide arrays ("GeneChips"). The former method originated with the advent of modern molecular biology, but the latter is a recently developed technique.

Transcriptional Fusion Libraries

One of the most commonly used methods by which to study gene expression is the construction of a transcriptional gene fusion (Sohasky *et.al.*, 1992). By joining the promoter (and controlling elements) of one gene to the structural component(s) of another gene (encoding for a detectable gene product), one may create a system by which gene expression can be assessed (Silhavy and Beckwith, 1985) (**Figure 1**). Referred to as "reporter genes", these structural components must encode for an obvious, or detectable, product. Three examples of commonly used reporter genes are *lacZ* (encoding for the enzyme β-galactosidase) (Casadaban, 1976), *luxAB* (encoding for the α and β subunits of

Figure 1. A Schematic of Transcriptional and Translation Gene Fusion Systems.

The first illustration indicates transcription and translation of gene X. If gene X is disrupted by a reporter gene (*eg. lacZ*), then a detectable product will be obtained. In the second system illustrated, the reporter gene possesses its own ribosomal binding site sequence, therefore a functional reporter gene product will be produced. In contrast, as in the third system pictured, a reporter gene lacking its own ribosomal binding sequence will lack signals for translation. Thus, a hybrid protein product is formed only if the reporter gene is fused in frame with respect to the start site of the target gene. RBS=ribosome binding site; stop=translational stop site; ATG=translational start site.


translation
transcription
promoter
RBS
ATG
stop
transcription
and translation
protein X
promoter
RBS
ATG
stop
Reporter
gene
transcription
and translation
Reporter enzyme eg. beta-galactosidase
promoter
RBS
ATG
stop
Reporter
gene
transcription
and translation
Hybrid protein product eg. beta-galactosidase

the enzyme luciferase) (Engebrecht *et.al.*, 1985), and *gfp* (encoding for green fluorescent protein) (Hastings, 1996). Moreover, the selected reporter gene must somehow be introduced into the strain of interest, often using transposons (Simon *et.al.*, 1989), or transposable bacteriophages (Casadaban and Cohen, 1979). (to be discussed later)

lacZ (β-galactosidase)

Since its discovery (Pardee *et.al.*, 1959), the regulation of the *lac* operon in *E. coli* has been well characterized. Therefore, this operon makes an ideal candidate as a reporter gene system. In addition, the substrate of this system is a sugar (lactose), and may be used as the sole carbon source in differential and selective media for the detection of fusion gene products (Silhavy and Beckwith, 1985). Since many bacterial strains are able to ferment lactose, one may opt to either delete the existing *lac* genes, or select a strain that is Δ*lac*, to ensure a "clean genetic background" on which to screen for *lacZ* expression. In other words, the initial *lacZ* expression must be minimized so that the gene fusion products will be more easily detectable. Furthermore, several different compounds are commercially available that can act within the system to induce, or detect the activity of, the *lac* operon, such as the compounds isopropyl-β-D-thiogalactosidase (IPTG), 5-bromo-4-chloro-3-indoyl-β-D-galactosidase (X-gal), and *o*-nitrophenyl-β-galactopyranoside (ONPG) (Oehler *et.al.*, 1990).

The *lacZ* gene encodes for an enzyme (β-galactosidase) that normally acts to cleave lactose within the cell. For the study of the expression of a *lacZ* gene that has been fused to a promoter, one can exploit the ability of β-galactosidase to cleave an analog of

lactose (X-gal) to yield a blue colour (Silhavy and Beckwith, 1985). Even in the absence of *lacY* (encoding for the enzyme lactose permease), the blue colour resulting from X-gal cleavage can be detected on solid media. Alternately, β-galactosidase can be assayed in liquid using a substrate which yields a yellow colour when hydrolyzed by the enzyme (ONPG) (Castilho *et.al.*, 1984).

By constructing transcriptional fusion libraries using *lacZ*, one can easily assay for gene expression. Assays for β-galactosidase production are sensitive, and the screening for *lac* mutants is both convenient and simple (Casadaban and Cohen, 1979). One slight disadvantage to the use of a *lacZ* reporter gene, though, is the possible masking of its expression by the presence of bacterial pigments. Several other reporter genes, such as *luxAB* and *gfp*, are able to circumvent this potential problem (to be discussed later).

***luxAB* (luciferase)**

In contrast to the expresion of a *lacZ* reporter gene, *luxAB* expression (*ie*. light emission) can be measured at extremely low levels with minimal endogenous background. Also, the emission of light can be localized (Sohasky *et.al.*, 1992). Numerous and diverse organisms are able to emit light, such as bacteria, fireflies, and fish, and bioluminescence genes do not appear to be evolutionarily conserved (Hastings, 1996). Several of these bioluminescent genes make exemplar reporter genes, having been cloned and functionally characterized. Furthermore, these genes often provide a "real-time" measurement of gene expression using a protein that is not native to the organism under study (Schauer *et.al.*, 1988). For example, the luminescence genes of *Vibrio harveyi* and *V. fischeri* have been

examined and exploited as reporter genes in a number of bacterial systems (Sohasky *et.al.*, 1992; Engebrecht *et.al.*, 1983).

Several different genes encode for the regulation and biochemical reactions required for bioluminescence in the bacteria *V. fischeri* and *V. harveyi* (Engebrecht *et.al.*, 1985). The genes involved include *luxA* and *luxB*, which encode the subunits of the enzyme luciferase. Being a mixed function oxidase, luciferase requires oxygen and a reduced flavin mononucleotide to produce light at a wavelength of 490 nm by the oxidation of the protein luciferin (Engebrecht *et.al.*, 1983). By the exogenous provision of a linear aldehyde to viable cells, the need to include the additional genes *luxCDE* within a gene fusion can be circumvented (Engebrecht *et.al.*, 1983).

Various means by which to measure the *luxAB* gene product exist, including exposure to photographic film, a luminometer, visual examination, or the use of a scintillation counter. Of note is the fact that light emission is not effectively masked by the production of cell pigments, unlike some other colorimetric reporter gene products (*ie. lacZ*) (Sohasky *et.al.*, 1992). In addition, it is essential that consistent amounts of cells must be used for each inoculation to ensure validity of results, since the cell density is proportional to the intensity of light emitted (Wolk *et.al.*, 1991). Multiple samples must also be taken to ensure a mean value for light emission (Sohasky *et.al.*, 1992).

gfp (Green Fluorescent Protein)

A second example of a bioluminescent reporter gene is that encoding for the green fluorescent protein (GFP) of the jellyfish *Aequorea victoria.* Green light is emitted from

this organism upon mechanical stimulation (Prasher, 1995), but when calcium binds to the photoprotein aequorin (found within *A. victoria*) *in vitro*, blue light is emitted (Chalfie *et.al.*, 1994). Within the jellyfish itself, green light (and not blue light) is produced due to a second protein (GFP) that derives its excitation energy from aequorin. Purified GFP is able to absorb blue light at 395 nm, and emit green light at a peak of 509 nm (Chalfie *et.al.*, 1994), with a shift in light emission wavelength resulting not from an external cofactor, but due to an intrinsic *p*-hydroxybenzylideneimidazolidinone chromophore that is generated by the cyclic oxidation of a Serine-Tyrosine-Glycine sequence (Heim *et.al.*, 1994).

The fluorescence of GFP does not appear to be species-specific, does not require any peculiar cofactors, the protein is quite small (27 kD), and can be monitored non-invasively. Therefore, GFP makes a suitable fluorescent reporter gene for the visualization of gene expression and cellular proteins (Stearns, 1995). In addition, GFP may be mutagenized to emit altered spectra, allowing for *in vivo* visualization of differential gene expression (Heim *et.al.*, 1994). Some altered GFP expression systems have recently become commercially available, and have proven to be extremely useful in such diverse systems as *E. coli*, *Drosophila*, yeast, and mammalian cells (Wang and Hazelrigg, 1994).

Construction of a transcriptional fusion

The question remains: how does one insert a reporter gene into the bacterial strain of interest? Most frequently, transposable elements are used to perform this task. Over

the past few decades, transposable elements have become useful for insertional mutagenesis, the introduction of selectable markers and portable restriction sites, and for carrying reporter genes (Simon *et.al.*, 1989). Since transposable elements possess a broad range in size, structure, and specificity, they provide an excellent mode of reporter gene transport for the construction of transcriptional fusions.

Transposons are short DNA segments that are capable of inserting into a host genome (Neidhardt *et.al.*, 1996). They are similar to insertion sequences (the simplest of the transposable elements), but carry additional genetic traits, such as antibiotic resistance or virulence factors. On the other hand, transposable bacteriophages carry few additional traits, but it is advantageous for the transport vehicle of a reporter gene to carry a selectable trait, such as antibiotic resistance, so that one can select for successful insertions. Several transposons (*eg.* Tn*3*, Tn*5*, Tn*10*), and transposable phages are useful within strains of *E. coli* and *Salmonella typhimurium* (with only certain strains of *S. typhimurium* being sensitive to Mu), whereas other transposable elements have been designed for use within other bacterial genera (Bremer *et.al.*, 1985).

Mu dl

The transcriptional fusion vehicle Mu dl carries an ampicillin resistance gene, and a *lacZ* reporter gene (Casadaban and Cohen, 1979). This 37 kilobase pair transposable element was constructed as a derivative of coliphage Mu, with its essential genes intact (*ie.* transposase and replication genes) (O'Connor and Malamy, 1983). To its merit, Mu dl carries a promoterless *lacZ* gene whose transcription can be used to monitor promoter

activity. In addition, the random insertion of Mu d1 into a host chromosome can be selected for in a single-step due to its transposon/prophage characteristics and the presence of the *bla* gene (Silhavy and Beckwith, 1985). By selecting for ampicillin resistance, one can create *lacZ* fusions with relative ease (Casadaban and Cohen, 1979).

Unfortunately, Mu d phages are competent for transposition and are consequently relatively unstable. Also, the Mu phage remains integrated next to the gene fusion it has created, thereby dictating that lysogens be maintained at a temperature of 32° C (due to a temperature sensitive Mu *c* repressor) (O'Connor and Malamy, 1983). Reacting to this problem, a method was devised whereby the Mu d prophage would be homologously recombined with a lambda phage to stabilize its integration into the host chromosome (Bremer *et.al.*, 1984). The new fusion vehicle, that need not be maintained at 32°C, was referred to as λp*lac*Mu.

λp*lac*Mu

This plaque-forming lambda phage was isolated by Bremer *et.al.* (1984), and can be used to isolate *lacZ* protein fusions. The derived phage carries Mu attachment sites (and can therefore randomly transpose), but maintains relative stability (as compared with Mu) due to its lambda origin. Therefore, λp*lac*Mu can withstand genetic manipulations, such as P1 transduction, without the gene fusion being disrupted. Thus, cloning of the

lacZ disrupted genes can be facilitated by the use of this transposable element (Bremer *et.al.*, 1984).

Transposons

An alternative to the use of the aforementioned transposable bacteriophage derivatives are the transposons. Like Mu d1, transposons do not move in a completely random fashion (Neidhardt *et.al.*, 1996). For instance, Tn*3* prefers plasmid targets, whereas Tn*5* has a predilection for chromosomal targets. To their advantage, transposons usually carry antibiotic resistance genes (Neidhardt *et.al.*, 1996). When selecting a transposon for the transport of a reporter gene, one must clarify the problem to be addressed and select a transposable element with an appropriate target and antibiotic resistance.

A number of broad host range transposons have been designed for the generation of transcriptional fusions within various bacterial genera, such as *Streptomyces* (Schauer *et.al*, 1988) and *Salmonella* (Niedhardt, 1996). For example, the most commonly used transposon for the engineering of *E.coli* is Tn*5*, since it transposes with high frequency within numerous Gram negative bacterial species, including non-enteric bacteria such as *Pseudomonas*, inserts randomly into target sequences, and demonstrates high stability upon insertion into the host genome (Simon *et.al.*, 1989). Variants of Tn*5* have been designed that carry new antibiotic resistance markers (*eg.* gentamycin resistance, and streptomycin resistance—two relatively uncommonly used antibiotics), and *lacZ* or *luxAB*

reporter genes. In consequence, one can usually select a transposon with appropriate antibiotic resistance, and the desired reporter gene.

Thus, reporter genes (*eg. lacZ*, *luxAB*, and *gfp*) may be introduced into diverse bacterial species using several different methodologies. By tailoring the reporter gene and its carrier to accommodate requirements for host range, and selectable markers, one may design a suitable experiment. The following are examples illustrating the applications of transcriptional fusion libraries.

Examples of transcriptional fusion libraries

lacZ

In an article published in 1996, A.T. Ulijasz and colleages wished to examine the means by which vancomycin resistance is induced within the bacterium *Enterococcus faecium*. Since vancomycin resistance in *E. faecium* is inducible by the expression of two coordinate operons (*vanHAX* and *vanRS*), a system was needed whereby one would be able to measure the induction of vancomycin resistance using antibiotics that act in a similar fashion (*ie*. by inhibiting peptidoglycan synthesis and cell wall hydrolytic enzymes). To accomplish this, the *lacZ* gene was placed under the control of the *vanH* promoter. This promoter controls inducible vancomycin resistance, and its expression is regulated by the *vanRS* operon of the same bacterial species.

First, a *vanH::lacZ* gene fusion was transferred from an *E.coli* strain to the non-pathogenic species *Bacillus subtilis*, using homologous recombination via transformation. *B. subtilis* was selected as the subject of study since it does not possess the vancomycin-inducible regulon that *E. faecium* does, nor does it contribute to the existing pool of antibiotic-resistant pathogenic bacteria that already exist within clinical settings. Once it had been demonstrated that this gene fusion was able to express *lacZ* (as indicated by blue colonies on plates containing X-gal as an indicator) upon induction with various antibiotics, the induction of *vanH* (*ie.* β-galactosidase production) was measured with respect to antibiotic concentration. It was noted that β-galactosidase production increased upon induction with antibiotics that inhibited cell wall synthesis. In addition, the presence of strains expressing *vanRS* also seemed to induce β-galactosidase production. It was thus concluded that the vancomycin resistance system is induced mostly by inhibitors of early cell wall synthesis (such as fosfomycin), and that the *lacZ*-based reporter system that had been created shows promise for future studies into gene expression with respect to antibiotic resistance (Ulijasz *et.al.*, 1996).

A second example of the use of a *lacZ* reporter gene system is the construction and screening of a *lacZ* transcriptional fusion library within *Salmonella typhimurium* (Ahmer *et.al.*, 1998). A *Salmonella* homologue of the *E. coli* gene *sdiA*, which is involved in quorum sensing, was cloned into an expression vector under the control of the arabinose-inducible P_{BAD} promoter. Initial disruption of the *sdiA* chromosomal gene by mutagenizing the selected strain with a transposon (Tn*3*), ensured that the gene would not be expressed at background levels. Secondly, the plasmid construct possessing *sdiA* under

P_{BAD} control was transformed into the test strain, thereby yielding a *S. typhimurium* strain with arabinose-dependent *sdiA* expression.

The plasmid-containing strain was then lysogenized using Mu dJ, a bacteriophage derivative with similar properties to Mu d1, and 10000 random *lacZY* transcriptional fusions were screened for SdiA responsiveness. By inducing transcription from P_{BAD} with arabinose, and repressing its transcription with glucose, ten fusions demonstrating positive regulation by SdiA were isolated and selected for further study. The Mu dJ-interrupted genes were subcloned, sequenced, and identified as SdiA-regulated genes. Thus, *lacZ* transcriptional fusions are useful for the construction and screening of libraries to identify the target genes of purported regulatory proteins.

luxAB

Since luciferase genes encode for a product that is not often native to the bacterial species in question, it makes a useful reporter gene. For instance, C.P. Wolk and colleagues used a Tn*5*-derived transposon to generate *luxAB* fusions within the *Anabaena* genome (1991). *Anabaena* is a filamentous cyanobacterium that forms heterocysts when deprived of nitrogen. Thus, it was hoped that the expression of genes responding to a decrease in nitrogen, or other environmental fluctuations, could be identified by observing the luminescence of transposed colonies.

Luminescent colonies resulting from the successful transposition of the *luxAB* genes were visualized using x-ray film, and subsequently exposed to an altered environment for a fixed amount of time. Those colonies demonstrating a differential

amount of light emission were selected for cloning and further study (Wolk *et.al.*, 1991). The advantage of using the *luxAB* genes within this system include the swift dispersal of the product (*ie*. light) to show decreases and increases in gene expression, and the sensitivity of the assay. Furthermore, since light is not obscured by cell pigments, which are common to cyanobacteria, luciferase activity is readily observable.

Aside from searching for inducible, or repressible, genes, one may use *luxAB* fusions to examine suspected transcriptional regulation. For example, a chromosomally located *E. coli* operon, that was first identified by screening a Mu d1 fusion library for luciferase production, has demonstrated involvement in the process of arsenic detoxification (Cai and DuBow, 1996). To further illustrate the transcriptional regulation of the genes involved in this operon, transcriptional fusions to these genes were created using the *luxAB* genes of *V. harveyi* (Guzzo and DuBow, 1994). By measuring the expression of these fusions (using a luminometer), it was noted that the operon was most inducible by sodium arsenite. In contrast, one gene product encoded by the operon was noted to have a negative effect on light emission. Thus, the regulation of this proposed detoxification operon was elucidated using the *luxAB* reporter genes.

gfp

Green fluorescent protein (GFP) is the "youngest" of the commonly used reporter genes, and has demonstrated use in tagging genes and cell structures (Prasher, 1995). One example of its usefulness is a study by L. Kremer and colleagues (1995), in which *gfp* (encoding for GFP) was placed under the control of the *hsp60* (heat shock) promoter from

BCG (*Mycobacterium bovis*) in *M. smegmatis*, as an expression marker. Since most other reporter genes require cell lysis and additional substrates, the mycobacteria are often killed before their gene expression has been effectively measured *in vivo*. In contrast, GFP allows expression to be monitored without the addition of substrates, and may be observed *in vivo* and in real time using fluorescence microscopy. Thus, the expression of *gfp*. under the control of *hsp60*, within BCG can be studied in culture, and within infected mice (Kremer *et.al.*, 1995).

By examining the induction of *gfp* expression by several different antibiotics, luminescence of was used to assess drug resistance within target cells. In addition, GFP was used to directly observe bacilli within infected macrophages. It was thus concluded that GFP may serve as a useful fluorescent indicator of gene expression in mycobacteria, both in tissue culture and *in vivo*. Using cytofluorimetry and by quantifying expression using a spectrofluorimeter, the usefulness of GFP in the study of pathogenic species of mycobacteria was confirmed.

Another example illustrating the use of GFP is a study in which cell-specific gene expression was visualized during sporulation in *B. subtilis* (Webb *et.al.*, 1995). Since sporulation is governed by several different RNA polymerase sigma factors, the *gfp* gene was placed under the control of four different promoters that are subject to three such sigma factors. Thus, the expression of each of these four promoters was visualized via *gfp* expression at different stages during sporulation.

Transcription of *gfp* (as visualized using phase-contrast microscopy) was induced by the various sigma factors specific to each of the four selected promoters. As expected, those promoters under the control of each specific sigma factor demonstrated localized,

cell-specific patterns of fluorescence (Webb *et.al.*, 1995). It was thus concluded that GFP is extremely helpful for the visualization of gene expression, and the localization of subcellular proteins, in *B. subtilis*. Since there are no exogenous substrates required (unlike luciferase), and cells need not be lysed or chemically fixed (unlike immunostaining with fluorescent antibodies), GFP can be used for *in vivo* studies of bacterial gene expression.

After examining several different studies, involving several different bacterial genera, it has been effectively illustrated that the construction of transcriptional gene fusions remains a beneficial tool in the study of gene expression (**Table 1**). Each of the above-mentioned reporter genes (*ie. lacZ, luxAB,* and *gfp*) will continue to be of tremendous assistance in the quest to understand prokaryotic gene expression and regulation.

Table 1. Transcriptional Gene Fusions

Component	Advantages	Disadvantages	Selected References
1. Reporter gene			
lacZ	- inexpensive - sensitive	- may be masked by bacterial pigments	Silhavy and Beckwith, 1985
luxAB	- sensitive - real-time	- often not native to study organism	Sohasky *et.al.*, 1992 Engebrecht *et.al.*, 1985
gfp	- no exogenous substrates - sensitive - real-time	- no major disadvantages	Prasher, 1995 Chalfie *et.al.*, 1994
Gene fusion vector			
Mu dl	- one-step gene fusions - random	- competent for transposition	Casadaban and Cohen, 1979
λp*lac*Mu	- excision can be selected	- **more** stable than Mu dl	Bremer *et.al.*, 1984
Transposons	- high frequency transposition numerous	- can be not entirely random	Simon *et.al.*, 1989 Niedhardt *et.al.*, 1996

Two-Dimensional Protein Electrophoresis Gels

Two-dimensional gels have proven invaluable for determining sets of genes induced by certain stimuli (O'Farrell, 1975). By isoelectric focussing in one dimension, and then running a conventional sodium dodecyl sulfate gel in a different dimension, a high resolution of proteins can be obtained (O'Farrell, 1977). The system of 2D polyacrylamide electrophoresis (2D PAGE) allows for the separation of heterogeneous mixtures of proteins at one time. For example, the proteins involved in the heat shock response of *E. coli* were identified using 2D PAGE (Lemaux *et.al.*, 1978).

For years after the development of 2D PAGE, the problem of standardization existed. To eliminate this quandary, the first set of reference gels was developed in 1980 (Bloch *et.al.*). Each reference gel provided a grid that assigned unique x and y axis numbers to each protein. This allowed for identification of those proteins matching the naming system on the reference gels.

At present, two different gene-protein databases exist whereby both genome expression and response regulation in *E.coli* have been catalogued (VonBogelen *et.al.*, 1996). Therefore, 2D PAGE will remain valuable in the quest to understand genetic responses to environmental stimuli, as well as the relationship between genome and protein expression.

Oligonucleotide arrays ("GeneChips")

One of the most recent advances towards DNA sequencing, and studies of genetic linkage, genetic variability, and gene expression, is the development of oligonucleotide arrays, or "GeneChips" (Fodor *et.al.*, 1991). "GeneChips" consist of high density arrays of DNA fragments, affixed to a small solid matrix (Johnston, 1998). By examining the hybridization of target molecules to the anchored oligonucleotide sequences, scientists can quickly assay samples against thousands of synthesized compounds.

While threatening the obsolescence of basic molecular biology techniques, the "GeneChip" has been revolutionizing the means by which scientists examine genes and their function. Since its recent conceptualization and design by the biotechnology company Affymetrix (Santa Clara, CA), this DNA microprobe has spawned numerous studies and modifications (Chee *et.al.*, 1996).

The Process

GeneChip technology entails several different methodologies. In essence, the techniques of photolithography and solid-phase chemical synthesis have been joined to create the initial steps in the process (Fodor *et.al.*, 1993). Once a chip has been exposed to target molecules, confocal fluorescence scanning enables the measurement of molecular binding events.

Photolithography is the technique by which semiconductors are produced (Fodor *et.al.*, 1991). By projecting ultraviolet light through a series of masks, multilayered circuits are etched into silicon chips. In a similar fashion, the GeneChip is created when light is used to direct the simultaneous synthesis of numerous chemical compounds upon a glass support. This is achieved by modifying chemical building blocks (such as deoxynucleosides) with photochemically removable protecting groups, and binding them to a solid substrate (Fodor, 1997). As light is conducted through a photolithographic mask onto certain areas on this surface, localized photodeprotection allows the chemical synthesis of compounds to occur. Light is then directed through a new mask, onto another region, and the chemical cycle repeats.

A general rule of the aforementioned chemical synthesis is that 4^n (where n=length) polynucleotides can be produced in $4xn$ chemical steps (Fodor *et.al.*, 1991). This seems to be a relatively low number when compared to the amount of data one can amass from a single array. In addition, the number of compounds per support has expanded rapidly since the conception of the GeneChip. From hundreds to thousands of compounds may now be created on a single oligonucleotide array (Lockhart *et.al.*, 1996). Undoubtedly, the sheer volume of data output far exceeds the magnitude of the chip's construction.

Once an array has been created, a sample of target molecules must be fluorescently labeled for subsequent detection using confocal fluorescence scanning. Firstly, the array is directly exposed to a solution of fluorescently labeled target molecules (*eg.* antibodies, receptors, single-stranded DNA or RNA). After allowing for hybridization, the chip is washed free of unbound molecules. A laser then excites target molecules, and fluorescence emission is collected by a microscope objective through a sensitive detector.

A two-dimensional fluorescence image can be obtained within minutes, and this image may be decoded using specialized GeneChip software (Fodor *et.al.*, 1993). Thus, by using oligonucleotide arrays one may dramatically decrease workload and time consumption, while inversely affecting the amount of data that is acquired.

Applications of the GeneChip

DNA Sequencing

DNA sequence analysis is the most obvious application for such oligonucleotide arrays (Pease *et.al.*, 1994). Conventional DNA sequencing is labour intensive and time consuming, whereas miniature oligonucleotide arrays enable parallel hybridization analysis that yields direct sequence information. This is referred to as "sequencing by hybridization" (SBH), where short probes of defined sequence are used to search for complementary sequences on longer target DNA (Sapolsky *et.al.*, 1996). Target DNA is added to the array, presenting probe of known sequence. The resultant hybridization pattern reveals the identification of all complementary probes. This method is extremely useful for genome sequencing projects (such as the Human Genome Project), and can be automated to maximize its efficiency.

Aside from primary DNA sequencing, SBH can be used to screen for base pair changes within a sequence (Pease *et.al.*, 1994). For example, oligonucleotide arrays are currently being used to screen the reverse transcriptase and protease genes of HIV-1 in the

quest to identify genetic variation and mutations conferring resistance to antiviral drugs (Lipshutz *et.al.*, 1995). By designing arrays that are capable of discerning differences between nucleic acid targets, so as to screen for changes, researchers have created a cost effective method for seeking mutations in those genes targeted by HIV-1 therapeutics.

Examining Genetic Polymorphisms

Mapping genomic libraries can pinpoint relevant chromosomal markers, and allow for the comparison of polymorphisms (Sapolsky *et.al.*, 1996). Thus, developmental and evolutionary processes may be clarified. With genetic markers lying adjacent to restriction endonuclease sites, oligonucleotide arrays can be synthesized so that they possess probes complementary to these markers. Therefore, the similarities between marker sets for various clones can be hybridized to these arrays, and their overlap can be measured. It is this overlap that could be used to construct a contig map of a number of different clones (Sapolsky *et.al.*, 1996). Instead of manipulating DNA sequencing gels, and functional studies, polymorphisms can be partially identified and characterized using this proposed method.

Studying Gene Expression

Regardless, how can the GeneChip interpret the relationship between primary DNA sequence and gene expression? This has been addressed by probing arrays with

cDNAs or mRNAs (Lockhart, 1996). Since only a fraction of genes encoded within a genome are expressed at a particular time, the parallel and simultaneous monitoring of genes seems only logical. Other methods exist for detecting mRNA, such as Northern blotting, but they are time-consuming, costly, and difficult to automate. Alternately, cDNA may be sequenced so that expressed genes can be identified. Unfortunately, this is not a very sensitive methodology. Thus, oligonucleotide arrays have been designed that are able to measure mRNA levels and, consequently, gene expression (Lockhart *et.al.*, 1996).

By measuring the hybridization of mRNA to pairs of 20-mer oligonucleotides with one being perfectly complementary to a subsequence of an mRNA message, and the other 20-mer being identical except for one base pair in its central position, it was found that the target cells are clearly identified (Lockhart et.al., 1996). These results indicate confidence in the ability of oligonucleotide arrays to unambiguously hybridize to target mRNA. Therefore, such arrays are efficacious for the study of gene expression.

Furthermore, bacterial mRNA transcripts have been studied by similar means (de Saizieu *et.al.*, 1998). A high-density oligonucleotide array was designed and created that presents probes depicting 106 genes from *Haemophilus influenzae*, and 100 genes from *Streptococcus pneumonie*. Using total RNA as a hybridization probe for the oligonucleotide array, concurrent quantification of transcript levels was achieved. Furthermore, upon comparison with Northern blot data attained for certain genes, the gene chip results correlated with respect to mRNA levels. Therefore, this methodology, along with the number of complete bacterial genomes being published, may be employed to understand bacterial gene expression. In particular, the study of virulence induction,

responses to environmental signals, and regulatory networks may be hastened and simplified. When compared to the above mentioned transcriptional fusion libraries constructed to study similar aspects of bacterial gene expression (Ulijasz et.al., 1996; Cai and DuBow, 1996), bacterial transcript imaging using oligonucleotide arrays seems both promising and feasible.

The Future

The development of the GeneChip has opened the portal to automated, quick, and efficient probes for DNA sequencing, polymorphism detection, pathogen identification, and gene expression monitoring (Johnston, 1998). Inevitably, the use of tried-and-true molecular biology protocols and methodologies (such as transcriptional fusion libraries and DNA sequencing) may be curbed. Ultimately, the study of gene expression will be reduced to a tiny microchip, effectively ushering in a "brave new world" in the field of functional genomics.

Introduction to the Ner-like protein (Nlp) of *Escherichia coli*

As previously mentioned, the relationship between genome and gene function is often not yet elucidated. A simple connection cannot be established without first understanding the levels of bacterial gene regulation. By constructing transcriptional fusion libraries, such as those using *lacZ*, *luxAB*, or *gfp* reporter genes, or by employing

GeneChips, one may clarify any indistinct correlations. For this thesis, a *lacZ* transcriptional fusion library was studied for its potential to examine the regulation of the *nlp* gene in *E. coli*.

Bacterial gene regulation often occurs at the transcriptional level. Transcription is one of the initial steps in gene expression. RNA polymerase sigma factors, bacteriophage repressors, and the *lac* repressor are familiar examples of regulatory proteins involved in the control of transcription. By transcriptional regulation, cells are able to alter gene expression to accommodate changes in growth conditions and exposure to environmental stresses. This genetic flexibility is what enables bacteria to grow under variable conditions.

A diverse array of DNA binding regulatory proteins, with some involved in transcriptional control, has been identified and characterized in both prokaryotes and eukaryotes (Pabo and Sauer, 1992). Several distinct DNA binding motifs exist, including the helix-turn-helix motif (as found in the lambda Cro protein, *E.coli* CRP, and the *lac* repressor), Zinc finger domains (such as some genes involved in *Drosophila* development), and leucine zippers (found in several eukaryotic transcription factors). These DNA binding proteins are categorized as such due to their related structural motifs for DNA recognition. A unique family of DNA binding proteins, with a conserved DNA binding domain, has emerged. This family is comprised of the negative early repressors (Ner) of temperate coliphages Mu and D108, the Ner-like protein (Nlp) of *Escherichia coli*, and the human TATA element modulatory factor (Tmf) (Autexier and DuBow, 1992).

The first members of this DNA binding protein family are Mu and D108 Ner. Mu and D108 are temperate coliphages capable of either lytic or lysogenic growth (Kukolj and DuBow, 1992). Both phages possess approximately 37kb linear double-stranded DNA genomes sharing 90% nucleotide sequence homology. During the lysogenic cycle, Mu and D108 integrate into the host genome at random locations, and are maintained by the *c* repressor (Kukolj and DuBow, 1992). The *c* repressor acts to inhibit transcription from the early promoter by binding to its three operator sites in Mu and two operator sites in D108, and consequently inhibiting the expression of early lytic genes (including Ner). On the other hand, during the lytic cycle, there is amplification and transposition of the phage genome initiated by the bacteriophage protein Ner. Ner, or negative early repressor, is the first gene expressed from the early promoter, and its life purpose is to bind to its binding site and inhibit transcription from the P_c promoter By binding to this site, Ner also inhibits its own production, as well as the expression of the early genes required for lytic growth. However, a host protein, integration host factor (IHF) binds to its own site and somehow (possibly by DNA bending) allows contortion of the DNA to enable RNA polymerase to transcribe the lytic genes. IHF is therefore essential for the expression of early genes (Kano *et.al.*, 1993).

The second, and most recently recognized, member of this protein family is of human origin. TATA element modulatory factor (Tmf) was first identified in 1992, by Joseph Garcia and colleagues, and was found to act as a transcriptional repressor capable of binding to the TATA element of HIV-1 long terminal repeat *in vitro*. By specifically binding to TATA elements (as determined using gel retardation assays), Tmf prevents TBP (TATA binding protein) from binding to TATA and therefore represses its basal

transcription of the HIV-1 long terminal repeat (Garcia *et.al.*, 1992). *Tmf* has been localized to chromosome 3p12-p21 and, interestingly, such changes in chromosome 3 have been associated with a number of different malignancies. Perhaps Tmf plays an important role in the regulation of viral and cellular gene expression, but this remains to be fully comprehended. It is interesting that both Tmf and Ner act as transcriptional repressors.

The third member of this unique DNA binding protein family is the Ner-like protein (Nlp) of *Escherichia coli* . Nlp was first identified by Yong-Lark Choi and colleagues in 1989 by its ability to stimulate maltose fermentation in a *crp*1* mutant of *E.coli*. A *crp*1* mutation encodes for an altered cAMP receptor protein (CRP). Recall that cAMP acts as a ligand by binding to CRP and allowing it to bind to a specific DNA promoter region to positively or negatively regulate transcription Normally, cAMP is required for activation of CRP, but *crp*1* most likely encodes for a mutant protein with a partially activated conformation (Aiba *et.al.*, 1985). CRP is involved in the activation of many sugar metabolism operons (in the absence of glucose), and this altered CRP results in reduced activation of the *mal* gene in the absence of cAMP. CRP*1 is, however, functional in the expression of most other sugar fermentation genes (including arabinose). CRP*1 is also able to activate expression of the *mal* operon in the presence of cGMP, therefore Choi and his colleagues wished to exploit this feature to isolate the putative gene encoding for guanylate cyclase (*cyg*). By transforming a library of cloned *E.coli* DNAs into the (*crp*1, cya-*) strain MK2001, and selecting for maltose fermentation (*ie.* a change from white to red when transformed with cloned DNA on plates containing maltose as their only fermentable sugar source, and a methyl red indicator), they found two clones.

One of these clones possessed a 1.7kb *E.coli* fragment that mapped to 69.3 minutes on the *E.coli* chromosome. This fragment was found to contain an open reading frame encoding for what was subsequently referred to as the *nlp* gene (due to its amino acid sequence homology with Ner of Mu and D108). The expression of the *nlp* gene is therefore able to complement a *crp*1* mutation and allow for maltose fermentation in a *crp*1* mutant strain of *E.coli* (Choi *et.al.*, 1989). In 1991, Kawamukai and colleagues identified several other genes that were able to stimulate maltose fermentation in MK2001. They named these genes *sfs* genes (for "sugar fermentation stimulation"), and henceforth referred to *nlp* as *sfs*7. For the purpose of this thesis, *sfs*7 will be referred to as *nlp*.

The *nlp* gene was found to encode for a 91 amino acid polypeptide of 10.4 kDa. The amino terminal region of Nlp was believed to contain its DNA binding region, and this 63 amino acid residue region is also well conserved between Mu and D108 Ner . Upon analysis, the DNA binding regions of Nlp and both Ner proteins, were found to be homologous to several known repressor proteins (Choi *et.al.*, 1989).

Nlp was thus named because it shares approximately 60% amino acid sequence homology with the Ner proteins of Mu and D108. Nlp shows a higher amino acid sequence homology to Mu Ner (63%) and D108 Ner (62%), than the Ner proteins share with each other (51%) (Benevides *et.al*, 1994). Since Tmf shares 44% amino acid sequence homology to Mu and D108 Ner in its DNA binding domain, it is considered to be part of the family. Interestingly, there also exits a Ner homologue within a cryptic prophage of *H.influenzae*.

It was determined that, despite its high degree of homology, Nlp cannot bind to the Mu or D108 Ner binding sites *in vitro*, nor does Nlp confer pseudoimmunity to either

phage *in vivo* (Autexier, 1991). Pseudoimmunity refers to the ability of overexpressed *ner* gene to act as a repressor and turn off Mu lytic functions in a host cell.

Through *nlp* insertions with *lac* and *luxAB* reporter genes, it was determined that *nlp* is expressed, but not essential for cell viability. A Southern Blot using DNA from several different bacterial genera illustrated that the *nlp* gene is highly conserved among members of *Enterobacteriaceae* (Autexier and DuBow, 1992). Hybridization of DNA containing the *nlp* gene was detected in several members of this family, with the highest homology to *Shigella sonnei*. *nlp* was also detected in some organisms distantly related to *E.coli*, such as *Vibrio harveyi* and *Pseudomonas aeruginosa*. Nlp may not be essential for cell viability, but the fact that it is conserved among many members of the *Enterobacteriaceae* indicates that it likely serves a functional role in the cell.

The binding sequence of Ner has shown homology to the binding sequence of cAMP receptor protein (CRP) and a model has been proposed for D108 and Mu Ner binding in which Ner bends DNA and contacts the operator at CRP-like sites with domains homologous to a helix-turn-helix motif. Therefore, since Nlp shares amino acid sequence homology with Ner, it was suspected that Nlp also acts as a DNA binding protein. To verify this, purified Nlp was used in electrophoretic mobility shift assays and shown to also bind to the CRP site. Thus, Nlp was categorized as a DNA binding regulatory protein (MacIntyre and DuBow, in preparation).

Once the Nlp protein had been partially characterized, further investigation into the gene encoding for this protein was pursued. Transcriptional analysis of *nlp* was performed using primer extensions and Northern blotting. It was thus determined that the transcriptional start site for *nlp* occurs 29 base pairs upstream of the ATG start codon and

that *nlp* encodes a monocistronic transcript of approximately 300 nucleotides (MacIntyre and DuBow, in preparation). Further evidence to support the claim that *nlp* is a monocistronic gene include the fact that there is a strong termination signal between *nlp* and its upstream gene *ispB* (isoprenoid biosynthesis) which is oriented in the same direction as *nlp*, and that its downstream gene (*murZ)* (which encodes in the opposite direction to *nlp*) is an essential gene in cell wall synthesis. It is interesting to note that neither Mu nor D108 *ner* are monocistronic genes.

Previous work examined when is *nlp* expressed during the growth curve. To determine the pattern of *nlp* expression during a growth curve of *E.coli*, a *nlp::lacZ* fusion was used. This fusion was constructed by inserting a *lac* promoter gene into the *E.coli* chromosome under the *nlp* promoter. By monitoring beta galactosidase production in liquid, the amount of *nlp* expression was determined (Fortin and DuBow, in preparation).

The expression of *nlp* was found to be maximal at early log phase, and then rapidly declines. This result suggests that perhaps Nlp is required for accelerated growth, or Nlp may assist in passage into the log phase of growth. Since *nlp* is not essential for cell viability (as illustrated by the aformentioned *nlp* disruptions), and since *nlp* is not constantly expressed throughout the growth curve, then it may be playing an accessory role in bacterial cell growth. Studies of what is regulating *nlp* expression may clarify its role.

It is presently acknowledged that *nlp* is an evolutionarily well-conserved, monocistronic gene encoding for a protein that has homologues in bacteria, bacteriophages and humans. It is also known that *nlp* is maximally expressed during the

early log phase of growth, but not essential for cell viability. Nonetheless, the function of Nlp (aside from the stimulation of maltose fermentation) remained a curiosity to be addressed.

During the course of this work, it was essential to develop a general means by which to determine what genes have promoters whose expression can be positively or negatively regulated by Nlp. To do this, a *lacZ* transcriptional fusion library was constructed and screened for its suitability in the future identification of Nlp function.

The first step in this process was to disrupt the *nlp* gene in the selected strain. If *nlp* is disrupted, then a plasmid-bound copy of the gene can be used to control its expression within the cell. Since *nlp* had been previously disrupted using the *luxAB* genes (Autexier, 1991), a *nlp::luxAB* fusion was placed into the wild type strain *E.coli* 40 (Δ *lac*) using a P1 lysate produced from the *nlp::luxAB* fusion (Seguin, 1995). These genes were transduced into *E.coli* 40, and then *nlp* on its chromosome was disrupted via recombination. Since this fusion also carries a tetracycline resistance marker, insertion was selected for on media containing this antibiotic.

Following P1 transduction of the genes of interest, tests to indicate the emission of light were used to determine whether the *luxAB* genes were transduced (and thus the *nlp* gene disrupted). It was found that the transduced strain emitted a significantly greater amount of light than the control strain. To further verify the successful transduction of these genes, Southern blotting was performed on total cellular DNA using a radioactively labeled 351 base pair fragment of the *nlp* gene. All evidence supported the successful construction of an *E.coli* 40 *nlp::luxAB*, tetR strain, henceforth referred to as LF20300.

Next, LF20300 was transformed with a plasmid that did not contain *nlp*. Vectors using the *ara* P_{BAD} promoter were decided upon, since this promoter can be tightly regulated and is not leaky like the commonly used *lac* promoter, and could be an effective tool in future studies (Guzman *et.al.*, 1995). In addition, control of the expression of the gene of interest can be obtained through the addition of certain sugars in the growth media.

L-arabinose metabolism is primarily under the control of the genes *ara* BAD, with AraC acting as both an activator and a repressor. In the presence of arabinose, AraC binds to this sugar, a conformational change ensues, and it binds to the activator site to enhance transcription from P_{BAD}. Alternately, a co-repressor (such as glucose) binds to AraC and causes a conformational change. Then this complex binds to the operator site and inhibits transcription from P_{BAD} (Guzman *et.al.*, 1995).

The plasmid used (pBAD18Cm) possesses a multiple cloning site under the control of the *ara* P_{BAD} promoter, and the plasmid may be selected for by its chloramphenicol resistance marker gene. The plasmid also possesses a gene encoding for the AraC regulatory protein, whereby arabinose would induce gene expression from P_{BAD}, and the addition of glucose would repress that expression.

To identify which genes were responsive to pBAD18Cm (and arabinose), a reporter gene was introduced at random locations throughout the chromosome of the strain of interest. To do this, LF20300 (the *nlp*-disrupted *E.coli* 40) containing pBAD18Cm was lysogenized using a Mud1 phage.

Mud1*lac* contains the left-end regulatory region of Mu phage, a *lacZ* reporter gene, an ampicillin resistance marker, and a small amount of the early lytic genes

(O'Connor and Malamy, 1983). This phage randomly inserts into the host chromosome, and must be maintained at 32°C, since its *c* repressor is temperature sensitive. As mentioned earlier, *c* repressor acts to maintain integrated phage, therefore a temperature shift to 37°C will cause prophage induction and cell death.

With the *lacZ* gene randomly inserted into the host chromosome, any disrupted non-essential genes will express a *lacZ* fusion protein transcribed from the chromosomal gene's promoter. Expression of *lacZ* (*ie.* beta galactosidase) will result in either a blue or white colony on plates containing x-gal as an indicator (Casadaban and Cohen, 1979).

Lysogens resulting from the lysogenization of LF20300 with pBAD18Cm (LF203018) were selected for on TCMG plates containing tetracycline (for the *nlp* disruption), chloramphenicol (for the plasmid), and ampicillin (for Mud1) with 0.2% glucose (to repress expression from P_{BAD}). These lysogens were mastered onto TCMG plates containing the same antibiotics and either 0.2% glucose, or 0.2% arabinose.

Five thousand lysogens were screened and thirty potential clones were obtained. Of these clones, eight showed consistent colours when remastered onto media under the same conditions. These clones were selected for further study. Since the clones obtained are most likely arabinose- or glucose-responsive, examples of such genes are listed in **Table 2**.

Table 2. Arabinose- and Glucose- Inducible Genes

Arabinose	Gene	Location	Function
	araA	2'	L-arabinose isomerase
	araB	2'	ribulokinase
	araC	2'	regulatory protein for arabinose catabolism
	araD	2'	L-ribulose phosphate-4-epimerase
	araFGH	43'	arabinose transport
	araE	64'	arabinose transport

Glucose

Glucose is widely involved in bacterial metabolism. For example, glucose represses the genes involved in the citric acid cycle when used as a sole carbon source. Unlike most other inducible transporters, the glucose transporter is constitutively expressed and is capable of blocking some other carbon transporters. Furthermore, glucose catabolism can impede the synthesis of catabolic enzymes for many other carbon sources.

Chapter II: Materials and Methods

CHAPTER II: Materials and Methods

2.1 Bacterial Strains and Plasmids

The bacterial strains and plasmids used in this work are listed and described in **Tables 3** and **4**, respectively.

2.2 Media and Antibiotics

Selection was accomplished using various antibiotics at concentrations of 40 μg/ml for ampicillin (amp; Novopharm), 50 μg/mL for chloramphenicol (cam; Sigma Scientific), 50 μg/mL for kanamycin (kan; Boehringer Mannheim), and 10 μg/mL for tetracycline (tet; Boehringer Mannheim). The media used included LB broth [1% (w/v) tryptone, 0.5% (w/v) yeast extract, 1% (w/v) NaCl; pH 7.0], and TCMG [1% (w/v) BBL trypticase, 0.5% (w/v) NaCl, 0.85% (w/v) agar, 10mM $MgSO_4$]. Approximately 30 mL of sterilized agar was poured per plate.

2.3 Restriction Endonuclease Digestions

DNA digestions were performed using 3 units of restriction endonuclease per μg of DNA to be hydrolyzed. DNA was hydrolyzed in digestion buffer [6 mM Tris-HCl pH 7.5, 6 mM $MgCl_2$, 75 mM NaCl, and 6 mM β-mercaptoethanol]. Digestion reactions were mixed in 1.5 mL Eppendorf tubes and allowed to react in a 37° C water bath for 2-4

hours. Digested samples were subsequently examined on agarose gels, or extracted using equal volumes of phenol and chloroform plus subsequent ethanol precipitation. Following centrifugation at 12000 x g for 30 minutes (4° C), the pellet was dried, and resuspended in 1 x TE. All of the restriction enzymes used in these procedures were purchased from Pharmacia, as was the digestion buffer in which they were used.

2.4 Ligations

For each ligation reaction, a total of 1 μg of the appropriate DNA fragments was mixed with 2 μL of linker-ligation buffer [60 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 15 mM DTT, 1 mM spermidine, 0.75 mM ATP, and 50 μg/mL gelatin], and 1 unit of T4 DNA ligase (Gibco BRL). Each ligation mixture was incubated overnight at 11° C, heat inactivated at 65°C for 15 minutes, and then diluted two-fold before being used to transform competent *E. coli* cells.

Table 3. *Escherichia coli* Strains Used in this Work.

Strain	Characteristics	Origin
40	Δ*pro*AB-*lac*, *trp*8am, *rps*L	Bukhari and Metlay, 1973
JM109	*rec*A1, *sup*E44, *end*A1, *hsd*R17, *gyr*A96, *rel*A1 *thi* Δ(*lac-pro*AB)	Viera and Messing, 1987
LF 20300	40 *nlp*::*lnxAB*, tet^R	Seguin, 1995
LF20318	LF20300 containing the plasmid pBAD18Cm	this work
BU1381	Mud1 (amp, *lac*), Δpro *lac*, Mu_{cts} 62	Casadaban and Cohen, 1979
DH1	*rec*A1, *end*A1, *gyr*A96 *thi*, *hsd*RM^-, *sup*E44	Hanahan, 1983

LF203003,

LF20304, derivatives of LF20300

LF203007 lysogenized with Mud1

LF203011 phage from *E. coli* BU1381

LF203014 *demonstrating arabinose*

LF203016 *responsiveness*

LF203023

LF203029

Table 4. Plasmids Used in This Work.

Plasmid	Characteristics	Origin
pUC120	amp^R, *lac* P/O, cloning and sequencing vector	Viera and Messing, 1982
pBR322	amp^R, tet^R cloning vector	Sutcliffe, 1979
pBAD18Cm	Cm^R cloning vector, P_{BAD} promoter, *araC*	Guzman *et.al.*, 1995
pMC1871	protein fusion vector, *lacZ*, tet^R	Shapiro, *et.al.*, 1983

2.5.1 Preparation of Rubidium Chloride Competent Cells

To prepare competent *E. coli* cells, a fresh overnight was subcultured (1:20) into 100 mL of Luria-Bertani (LB) broth [1% tryptone, 1% NaCl, 0.5% Bacto yeast extract, and 2 mM NaOH]. The cells were allowed to grow in a 37° C shaking incubator until an absorbance (A_{550}) of 0.48 was reached. The cells were placed on ice for 5 minutes, and then centrifuged at 4000 x g for 15 minutes (4° C). The resulting cell pellet was resuspended in 40 mL of TFB I [30 mM CH_3COOK, 100 mmol $RbCl_2$, 10 mM $CaCl_2$-$2H_2O$, 50 mM $MnCl_2$-$4H_2O$, 15% glucose; pH 5.8] and placed on ice for 5 minutes. The cells were centrifuged at 4000 x g for 15 minutes (4° C). The cell pellet formed was resuspended in 4 mL of TFB II [10 mM MOPS, 75 mM $CaCl_2$-$2H_2O$, 10mM $RbCl_2$, 15% glycerol; pH 6.3], and left on ice for 15 minutes. This suspension was divided into 200 µL aliquots, quick-frozen in an ethanol dry ice bath, and stored at -70 °C.

2.5.2 Transformation using Rubidium Chloride Competent Cells

A 200 µL aliquot of rubidium chloride competent cells was thawed at room temperature, and subsequently placed on ice for 10 minutes. Less than 100 µg of DNA was added to the cells and mixed thoroughly. The cells were placed on ice for 30 minutes. Following this, the cells were heat shocked at 42° C for 90 seconds and then returned to ice for 1-2 minutes. LB (800 µL) was added to the suspension, which was then incubated

for 1 hour at 37° C. The cells were plated onto media containing the appropriate antibiotics for the selection of transformants, and incubated overnight at 37°C.

2.6 Construction and Screening of a *lacZ* fusion library (Figure 3)

2.6.1 Preparation of Mud1 (amp^R, lac) Bacteriophage Lysates

The Mud 1 containing *E. coli* strain BU1381 was streaked from a glycerol stock stored at -20°C onto LB agar plates containing ampicillin in order to obtain single colonies. By replica plating the obtained colonies onto a lawn of *E. coli* 40, it was determined whether they were capable of producing phage. The colonies that were able to lyse *E. coli* 40 cells were selected for production of the Mud1 lysate. Ten mL of LB broth containing ampicillin was inoculated with a single colony, and allowed to grow overnight at 32° C. This culture was diluted 50 fold in LB containing ampicillin, $MgSO_4$ (0.01 M), and $CaCl_2$ (0.005 M) and grown to early exponential phase (10^8 cells/mL). Following incubation, the cells were induced at 42° C for 30 minutes, then shifted to 37° C until lysis occurred (approximately 60 minutes). The culture was centrifuged for 8 minutes at 5000 x g (4° C) to obtain a lysate. This lysate was titered using *E. coli* 40. An *E. coli* culture was grown overnight at 32° C, diluted two-fold in LB, and grown for another 1 hour at 32° C. Two hundred µL of diluted *E. coli* culture were added to 5 mL test tubes each containing 2.5 mL of 0.5% soft agar, and vortexed for 3 seconds. One hundred µL of diluted Mud 1 lysate (10^{-3}, 10^{-5}, 10^{-7}, 10^{-9}) was added to the mixture and gently vortexed. The contents of the tube was quickly poured onto TCMG plates and

incubated overnight at 37° C. The resulting plaques were counted to determine the number of plaque- forming units per millilitre (PFU/mL) for the lysate.

2.6.2 Lysogenization of LF20318 Using a Mud1 Lysate

The recipient strain (LF20318) was grown overnight in a 10 mL culture containing tetracycline. The cells were diluted (1:20) in LB containing $MgSO_4$ (0.01 M), $CaCl_2$ (0.005 M), and tetracycline, and grown to early exponential phase at 32° C. At this stage, the Mu d1 bacteriophages were allowed to adsorb to the *E. coli* 40 cells for 20 minutes at 30° C at a multiplicity of infection (M.O.I.) of 0.1. Subsequently, 1.8 mL of LB was added to each mixture and incubated for 30 minutes at 30° C in a shaking incubator. The cells were then plated onto LB agar containing ampicillin, chloramphenicol and tetracycline, and incubated overnight at 32° C to select for ampicillin resistant lysogens. Five thousand lysogens were mastered onto TCMG plates.

2.6.3 Screening of the *lacZ* Fusion Library

To determine whether or not the genes interrupted by *lacZ* (via Mud 1) were arabinose responsive, the obtained lysogens were replica plated onto TCMG plates that would either induce [0.2% arabinose, amp, cam, tet, X-gal] or repress [0.2% glucose, amp, cam, tet, X-gal] expression from the P_{BAD} promoter of the plasmid pBAD18Cm. The lysogens were also mastered onto LB agar containing ampicillin, chloramphenicol, and tetracyline for the purpose of storage. Thirty lysogens demonstrated different *lacZ*

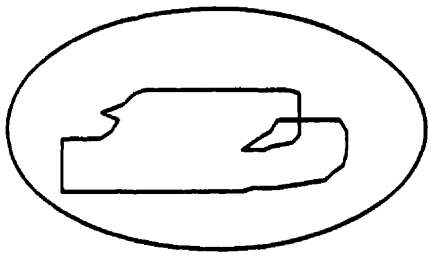
expression in the presence of arabinose, as compared to its absence. Of these potential arabinose-responsive clones, eight proved to be consistent for colour changes when repeatedly restreaked onto the same media. These 8 clones were further characterized and studied.

2.7 Isolation of Total Genomic DNA from Arabinose-Responsive Candidate Clones

Chromosomal DNA from the strains LF20303, LF20304, LF20307, LF20311, LF20314, LF20316, LF20323, and LF20329 was isolated. Overnight cell cultures of each strain were centrifuged in 50 mL conical tubes for 15 minutes at 5000 x g (4° C). The resulting cell pellets were resuspended in 1.4 mL of 10x TE. To each suspension, 200 µL of 10% (w/v) SDS and 200 µL of RNase A [1mg/mL in 10 mM Tris-HCl; pH 7.6] was added, mixed, and incubated at 37° C for 2 hours. Following incubation, 100 µL of pronase [20 mg/mL in 10 mM Tris-HCl; pH 7.6] was added, mixed, and allowed to incubate for 2 hours at 37° C. An equal volume of buffer saturated phenol (pH 7.0) was added, mixed by inversion, and centrifuged at 5000 x g (4° C) for 15 minutes. The upper aqueous phase was transferred to a new 14 mL Falcon tube, and two more phenol extractions were performed. The DNA was then extracted three times using an equal volume of ether. Ethanol precipitation was then performed, and the fibrous DNA pellet formed was removed from the solution using a 1-5 µL micropipette that had been sealed at one end. The DNA pellets were dried and resuspended in 50 µL of 1xTE.

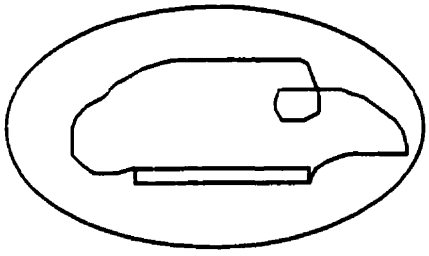
Figure 3. Construction of *lacZ* fusion library.

A *lacZ* fusion library was constructed by transforming the *E.coli* strain LF20300 (*nlp::luxAB*, tet^R) with the plasmid pBAD18Cm. The resultant strain was lysogenized with Mud1 *lac* (amp^R, *lac*), and lysogens were screened for *lacZ* expression on media inducing or repressing *lacZ* expression (*ie.* containing either 0.2% arabinose, or 0.2% glucose).



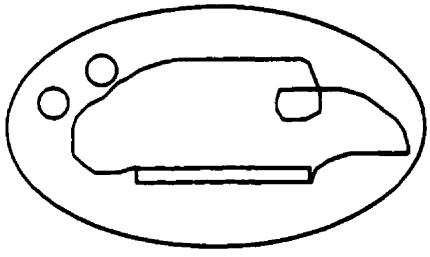
E. coli 40

P1 transduction of *nlp::luxAB* (tet^R)



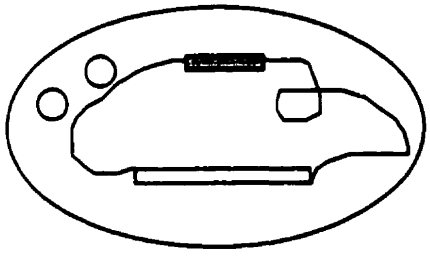
LF20300 (tet^R)

transformation with pSM10



LF20318 (tet^R, cam^R)

lysogenization with Mu d1 *lac* (amp^R)



screening on 0.2% glucose, or arabinose, plus X-gal

LF20303, and other Nlp-responsive clones (tet^R, cam^R, amp^R)

2.8 Southern Blotting to Determine Mud1 *lacZ* Copy Number and Fusion Restriction Maps

Total genomic DNA for each of the eight clones was digested with several restriction endonucleases, under previously described conditions. The enzymes used included *Bgl*II, *Bam*HI, *Eco*RI, *Eco*RV, and *Sac*I. The digested DNA fragments were electrophoresed overnight at approximately 25V on a 0.6% agarose gel along with uncut DNA and a molecular weight marker (lambda DNA hydrolyzed with *Hind*III). The gels were treated with 0.2 N HCl for 15 minutes, rinsed in deionized water, and then soaked in denaturing solution [1.5M NaCl/ 0.5 N NaOH] for 45 minutes. The gels were rinsed again in deionized water, and then soaked for 45 minutes in denaturing solution [1 MTris; pH 7.4/ 1.5 M NaCl]. After rinsing the gels again in deionized water, the DNA fragments were transferred to a nylon membrane (Hybond-N, Amersham) by capillary transfer in 20 x SSC [17.53% (w/v) NaCl, 8.82% (w/v) sodium citrate; pH 7.0] and allowed to run for 12 hours. The lanes for each gel were marked on their respective membranes, and the membranes were soaked in 6 x SSC for 5 minutes. The DNA was fixed using UV cross linking. Following cross-linking, the membranes were placed inside a bag and soaked in prehybridization solution [1 M NaH_2PO_4, 1 M Na_2HPO_4, 10% (w/v) BSA, 20% (w/v) SDS, 0.5 M EDTA, calf thymus DNA (10 mg/mL, and deionized water] for 1 hour at 68° A DNA fragment containing the *lacZ* gene from pMC1871 was labeled via random priming. Random priming was accomplished by adding the random prime mixture [10μL 0.2M KH_2PO_4;pH7, 4μL 50mM $MgCl_2$, 8μL cold dNTPs, 4μL radioactive dNTPs, 4μL hexanucleotide random primers] to the denatured DNA fragment plus 1.33μL 0.3M DTT, and 1μL Klenow fragment. The mixture was incubated for 3 hours at room temperature,

and stopped by the addition of 2μL 0.5M EDTA. Any non-incorporated nucleotides were eliminated by running the probe through a Sephadex G-50 column. The radioactivity of the purified probe was measured in a scintillation counter and found to be 1.5×10^7 cpm. After being denatured (via boiling), and added to the solution, the probe was allowed to hybridize overnight at 68° C in hybridization solution [4.3mL 1M NaH_2PO_4; 5.7mL 1M Na_2HPO_4; 2mL 10% BSA; 7mL 20% SDS; 40μL 0.5M EDTA; 200μL calf thymus DNA (10mg/mL); 950μL dH20]. Following incubation, the membrane was washed twice [6 x SSC, 0.1% SDS], and exposed to Kodak XAR X-Ray film inside a lead intensity screen for 24 hours at -20°C.

2.9 Identification of Candidate Arabinose-Reponsive Genes

Candidate arabinose-responsive genes were identified using putative restriction maps for each clone, as determined via Southern hybridization. By searching the entire restriction map of *E. coli* (Miller,1992) for a match to the determined maps, several genes were found to possess an identical restriction pattern. Those with orientation in the opposite direction to *lacZ* were not considered as candidate genes.

Chapter III: Results

Chapter III: Results

3.1 Plasmid pBAD18Cm (Figure 3)

After obtaining the plasmid pBAD18Cm, test digestions were performed to verify its sequence. The fragment sizes obtained correctly reflected the expected restriction endonuclease fragment pattern using the chosen enzymes. The plasmid pBADCm was used for subsequent experiments to determine its potential in the construction of gene fusion libraries.

3.2 Construction of *lacZ* Fusion Library

3.2.1 Preparation of Mu d1 Bacteriophage Lysate

A Mu d1 phage lysate was prepared using the strain BU1381, and titers between 10^6 and 10^{11} PFU/mL were obtained. These lysates were used at an M.O.I. of 0.1 to lysogenize the strain LF20318.

3.2.2 Lysogenization and Screening of LF20300 with Mu d1 Transducing Bacteriophage

The bacterial strain LF20318 was lysogenized with the transcriptional fusion vehicle Mu d1, and lysogens were selected for on TCMG media containing amp, cam, and tet. Five thousand colonies were selected for further testing by mastering on selective (via antibiotics) and differential (with 0.2% arabinose, or 0.2% glucose) media for *lacZ* expression **(Figure 4)**.

Figure 3. pBAD18Cm

The plasmid pBAD18Cm possesses the gene encoding for the AraC regulatory protein, the P_{BAD} promoter, and a multiple cloning site. Following the transformation of pBAD18Cm into the *E.coli* strain LF20300, the resultant strain was referred to as LF20318.

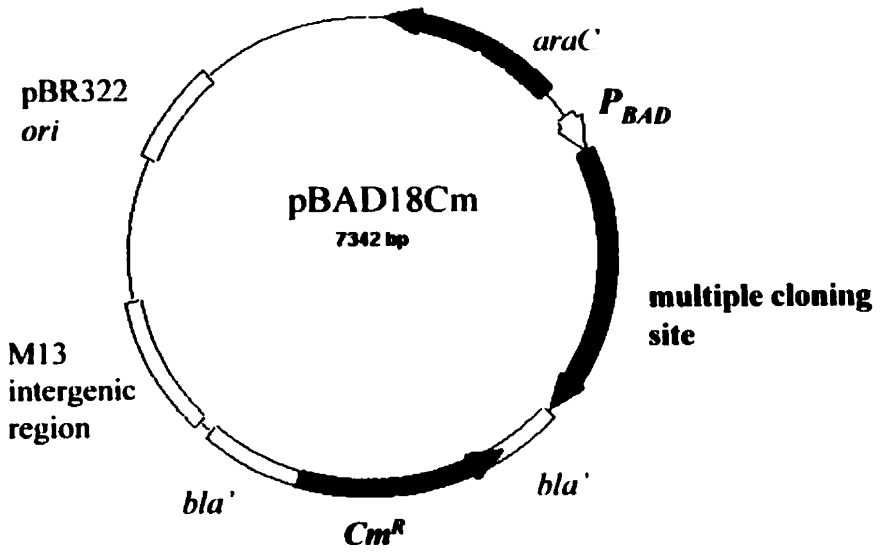
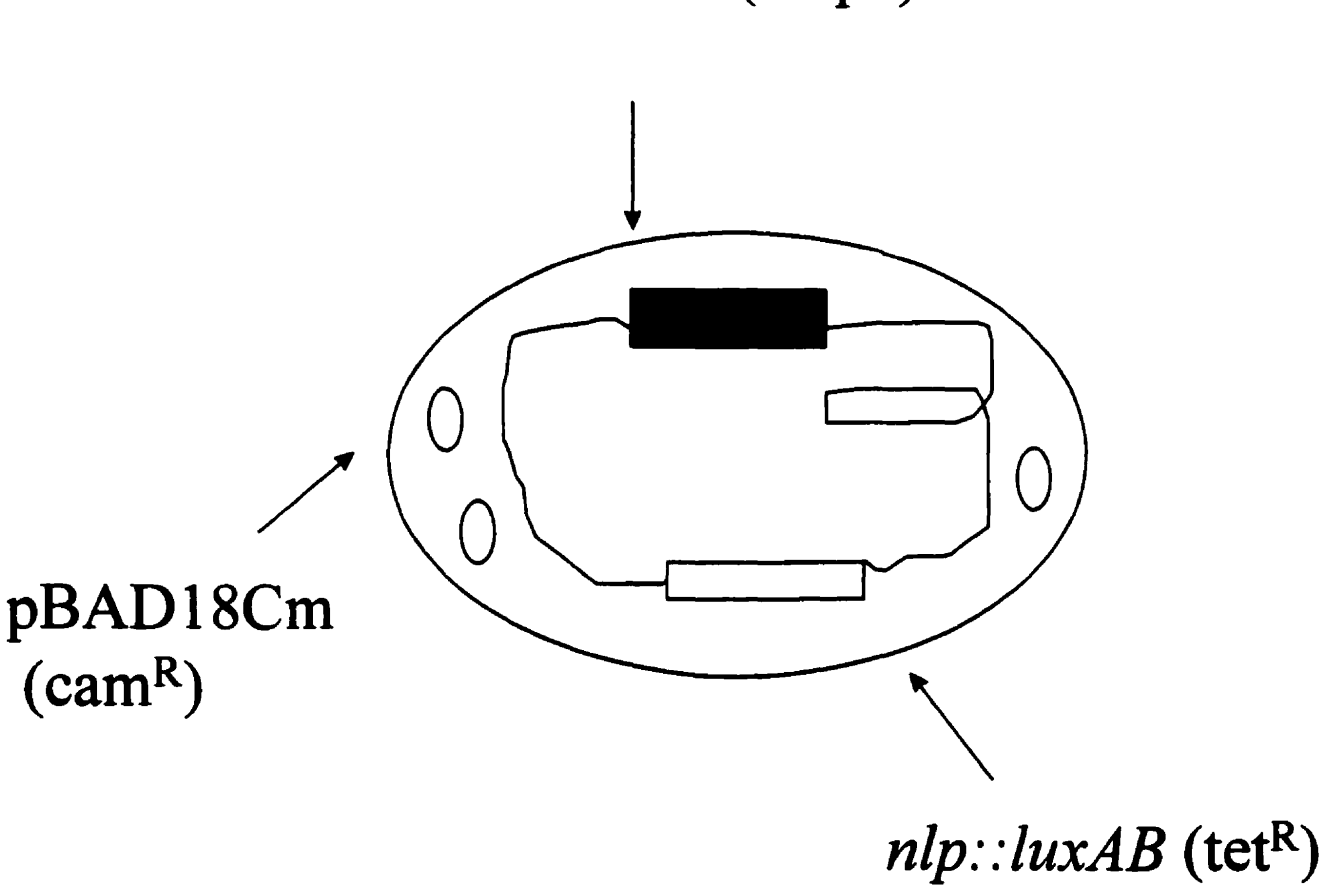

araC'
P_{BAD}
pBR322
ori
pBAD18Cm
7342 bp
multiple cloning
site
M13
intergenic
region
bla'
bla'
Cm^R

Figure 4. The System Used

The strain LF20318, lysogenized with Mu d1, is illustrated. Since no gene was placed under the control of the P_{BAD} promoter of this plasmid, the expression of *lacZ* should not be induced by the addition of arabinose unless the *lacZ* disrupted gene is arabinose-inducible. There are four possible outcomes in the presence of glucose or arabinose.

LF230318



plus glucose		plus arabinose	effect of arabinose
white	→	blue	+
blue	→	white	-
blue	→	blue	no effect
white	→	white	no effect

Table 5. Qualitative Analysis of *lacZ* Expression by Arabinose-responsive Clones

Arabinose-Responsive Clone	Colour in the Presence of Glucose	Colour in the Presence of Arabinose
LF20303	+ +	+ + + + +
LF20304	+ +	+ + + + +
LF20307	+	+ +
LF20311	+ +	+ + + + +
LF20314	+ +	+ + + + +
LF20316	+ +	+
LF20323	+	+ + + + +
LF20329	+ + + + +	+ +

Legend

+ = white

+ + = pale blue

+ + + = darker blue

+ + + + = blue

+ + + + + = dark blue

3.3 Determination of Mu d1 Copy Number and *lacZ* Fragment Size in Arabinose-Responsive Clones via Southern Blotting (Figure 5)

To determine the *lacZ* fragment sizes within each of the eight isolated clones, total genomic DNA was isolated and hydrolyzed with a number of different restriction endonucleases. DNA digested using the enzymes *Bam*HI, *Bgl*I, *Eco*RI, *Eco*RV, and *Sac*I, plus uncut genomic DNA, was electrophoresed overnight through a 0.6% agarose gel. The DNA was transferred to a nitrocellulose membrane, and probed using the *lacZ* gene radioactively labeled with ^{32}P. Using the known restriction map for Mu d1 **(Figure 6)**, the number of Mu d1 insertions, and the sizes of *lacZ* fragments were obtained. By observing the number of hybridization signals that appear for each digestion, the number of Mu d1 inserts was determined. Since *Bgl*II cuts in Mu d1 resulting in one hybridization signal, the number of bands for each clone digested with this enzyme was observed. LF20303, LF20304, LF20307, LF20311, and LF20314 demonstrated one hybridization signal with *lacZ* when digested with *Bgl*II, whereas LF20316, LF20323, and LF20329 demonstrated more than one. Therefore, it was concluded that LF20316, LF20323, and LF20329 likely possess two Mu d1 inserts. **Figures 7, 8, 9 and 10** illustrate the restriction maps for the eight clones, as determined by Southern blotting.

Figure 5. Southern Blot analysis of Arabinose-responsive clones (LF20303, and LF20329)

Total genomic DNA was isolated from each clone and subjected to hydrolyzation with several different restriction endonucleases. Examples of a clone containing a single Mu d1 insert (LF20303), and one containing multiple Mu d1 inserts (LF20329) are shown. U=uncut DNA; B=*Bam*HI; G=*Bgl*II; E=*Eco*RI; R=*Eco*RV; S=*Sac*I.

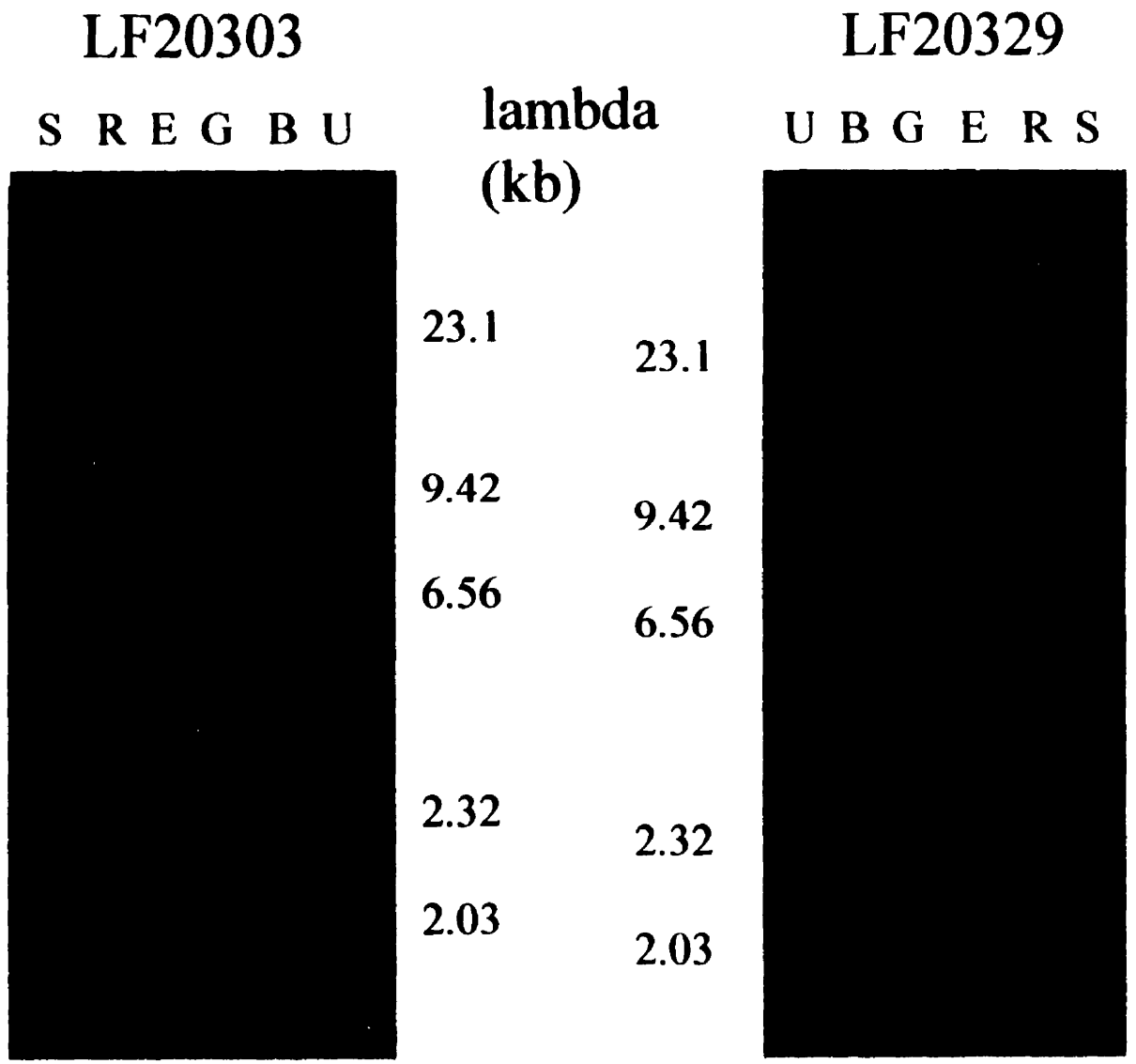
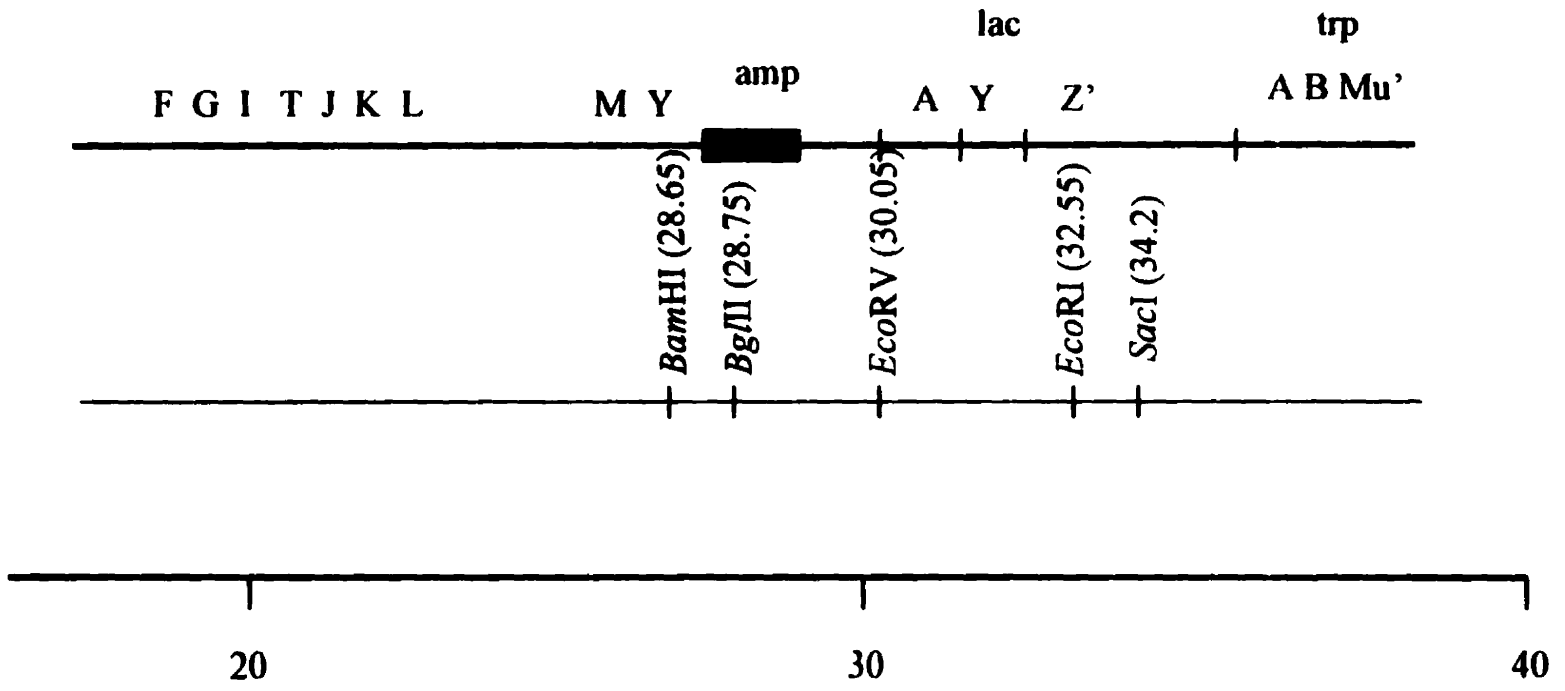

LF20303
S R E G B U
lambda (kb)
23.1
9.42
6.56
2.32
2.03
LF20329
U B G E R S
23.1
9.42
6.56
2.32
2.03

Figure 6. Restriction map of Mu d1 *lac* (O'Connor and Malamy, 1983)

The restriction sites contained within the Mud1 lac sequence were used, in conjunction with the Southern Blot analysis data, to deduce putative restriction maps for several clones.

Mud1 (amp, *lac*)

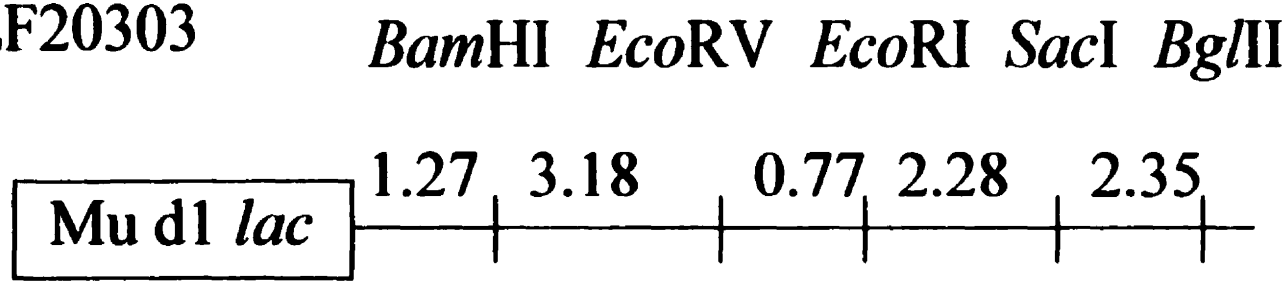


Kilobase pairs

Figure 7. Restriction maps of LF20303 and LF20304

The restriction sites around the Mud1 *lac* insertion within clones LF20303 and LF20304 are illustrated in Figure 7. Distances between each restriction site are indicated, and are measured in kilobase pairs.

LF20303



LF20304

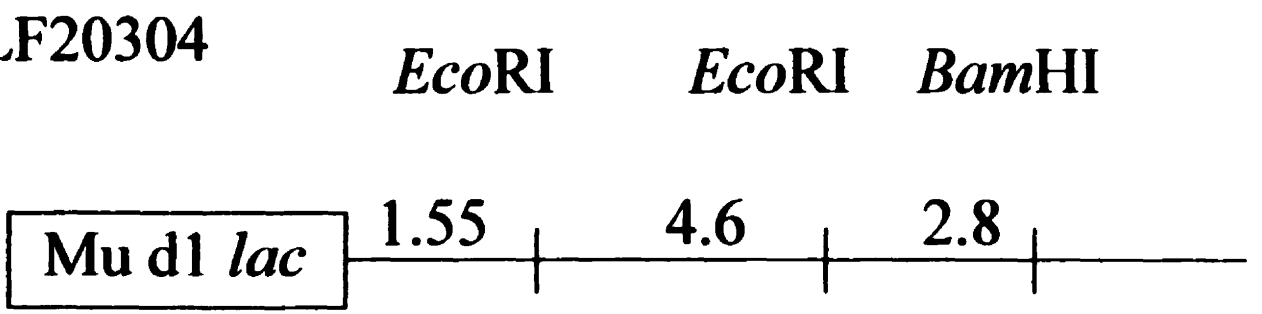
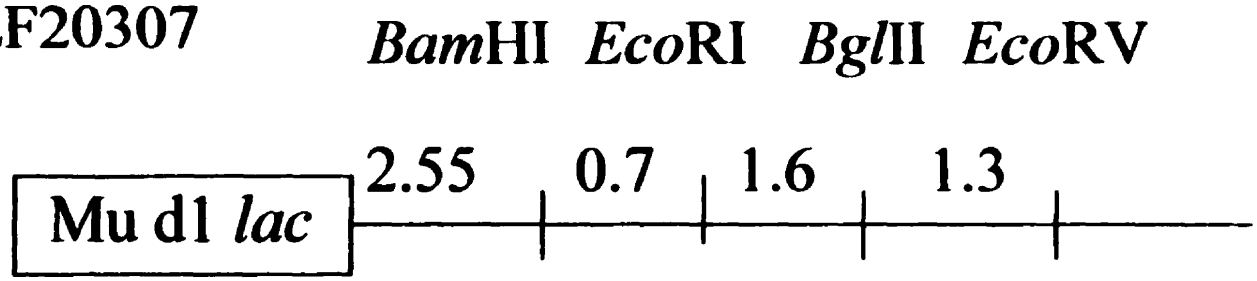


Figure 8. Restriction maps of LF20307 and LF20311

The restriction sites around the Mud1 *lac* insertion within clones LF20307 and LF20311 are illustrated in Figure 8. Distances between each restriction site are indicated, and are measured in kilobase pairs.

LF20307



LF20311

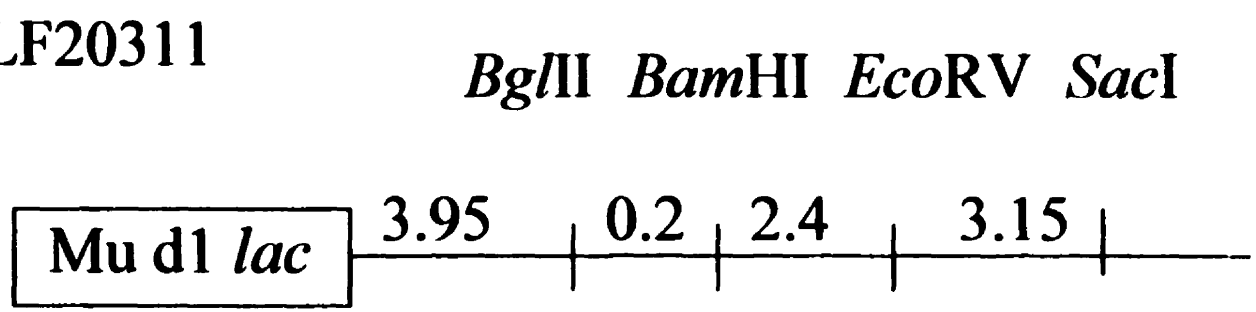
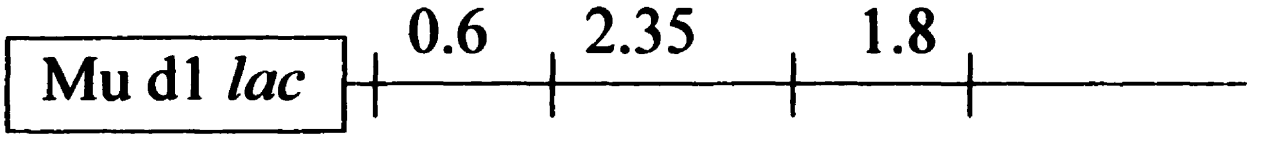


Figure 9. Restriction maps of LF20314 and LF20316

The restriction sites around the Mud1 *lac* insertion within clones LF20314 and LF20316 are illustrated in Figure 9. Distances between each restriction site are indicated, and are measured in kilobase pairs.

LF20314 *Bam*HI *Eco*RV *Eco*RI *Sac*I



LF20316 *Bam*HI *Eco*RV *Sac*I *Bgl*II

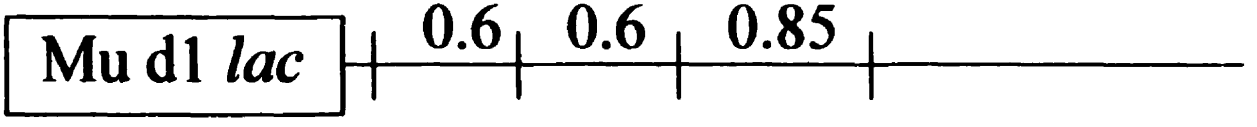
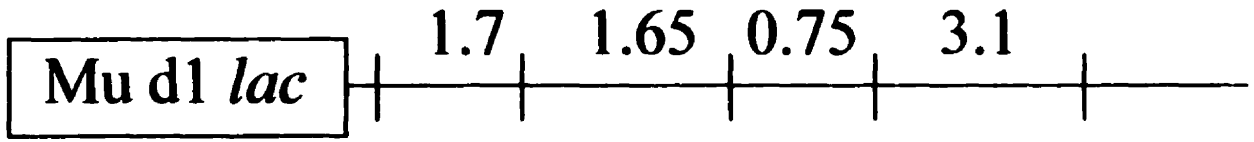


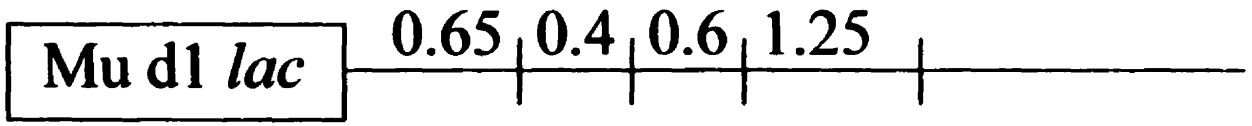
Figure 10. Restriction maps of LF20323 and LF20329

The restriction sites around the Mud1 *lac* insertion within clones LF20323 and LF20329 are illustrated in Figure 10. Distances between each restriction site are indicated, and are measured in kilobase pairs.

LF20323 *Bam*HI *Eco*RV *Bgl*II *Sac*I *Eco*RI



LF20329 *Bam*HI *Eco*RV *Bgl*II *Sac*I



3.5 Determination of Candidate Arabinose-Responsive Genes

Using the putative restriction maps of each clone, as determined by Southern blotting, the entire restriction map for *E. coli* was scanned for identical patterns. Several genes were identified for each clone, but those showing opposite orientation to *lacZ* were removed from the list. In addition, any essential genes were eliminated, since their disruption by *lacZ* would prevent cell viability. **Table 2** lists some arabinose- and glucose- inducible genes (*ie*. what was expected), and **Table 6** lists some of the obtained "candidate" genes.

Table 6. Candidate Genes

Clone	Gene	Location	Gene Function
LF20303	*betA*, *B*, or *C*	7'	choline metabolism and transport
	ubiX	50'	ubiquinone
LF20304	*ndh*	22'	respiratory NADH dehydrogenase
	dsdA, or *C*	51'	D-serine metabolism
LF20307	*purA*	95'	purine synthesis
	murH	99'	peptidoglycan synthesis
LF20311	*sad*	34'	succinate semialdehyde dehydrogenase
LF20314	*purE*, or *K*	12'	purine biosynthesis
	fadB	40'	fatty acid degradation
LF20316	*xapA*, or *R*	52'	xanthose metabolism
	aroE	72'	aromatic dehydroxylase
LF20323	*fruA*, *B*, *F*, or *K*	47'	fructose metabolism
LF20329	*cydC*	19'	cycloserine
	purA	95'	purine synthesis

Chapter IV: Discussion

Chapter IV: Discussion

After first being identified by its ability to stimulate maltose fermentation in the *E.coli* strain MK2001 (*cya, crp*1*) (Choi *et.al.*, 1989), Nlp's function within wild type strains remains unknown. It has been determined that *nlp* exists within several unrelated bacterial genera (using Southern hybridization), yet is not essential for cell viability (Autexier and DuBow, 1991).

The expression of *nlp* has demonstrated a maximum at entry into log phase using both *lacZ* and *luxAB* fusions (Fortin and DuBow, in preparation). Moreover, *nlp* expression is consistently lower throughout the rest of the log phase, and at the beginning of the stationary phase of growth. Perhaps Nlp is required for accelerated growth, or entry into the logarithmic phase. Interestingly, the histone-like protein Fis demonstrates a similar pattern of expression during the bacterial growth cycle (Ball *et al.*, 1992). It is also interesting that, like Nlp, histone-like proteins are not required for cell growth (Kano and Immamato, 1990). It is therefore possible that Nlp acts as a stimulator of sugar metabolism (its hypothesized role), or as a global regulator within wild type *E. coli*.

A putative helix-turn-helix motif has been identified in Nlp (Choi *et.al.*, 1989). This supports its potential function as a DNA binding protein, and perhaps a transcriptional regulator. To specifically determine the function of Nlp, an *E.coli* strain containing *nlp::luxAB* (tet^R) (LF20300) was constructed (Autexier and DuBow, 1991). This strain was subsequently transformed using a plasmid relying on the arabinose P_{BAD} promoter (pBAD18Cm). By using a vector without the cloned *nlp* gene, a preliminary study was performed to provide a control for future experiments examining the expression

of this gene. Since this plasmid demonstrates chloramphenicol resistance, successful transformants could be selected for on LB containing tet and cam. Next, the resultant strain (LF20318) was lysogenized using Mu d1 (amp, *lac*). Lysogens were selected on TCMG plates containing tet, cam, and amp. Five thousand colonies were mastered onto three sets of plates: TCMG containing tet, amp, cam, and X-gal plus either 0.2% glucose (to suppress expression from P_{BAD}), 0.2% arabinose (to induce expression from P_{BAD}), and LB containing tet, cam, and amp. Eight clones (of the 5000 mastered colonies) consistently demonstrated arabinose-responsiveness with respect to *lacZ* expression. These clones were henceforth referred to as LF20303, 04, 07, 11, 14, 16, 23, and 29. Of the eight arabinose-responsive clones, two turned from blue to white in the presence of arabinose, and the other six turned white to blue in the presence of arabinose. These clones were selected for further study. Since no gene was cloned into the multiple cloning site of pBAD18Cm, it was suspected that the identified genes would be regulated by these two sugars, and thus provide important data as to the background of *nlp*-responsive genes isolated by this procedure.

To ascertain the identity of the *lacZ*-disrupted genes, total genomic DNA was isolated from each of the eight clones, and digested with several different restriction endonucleases. The digestions were probed in a Southern hybridization using the *lacZ* fragment from pMC1871 (Figure 6). Putative restriction maps were determined for those clones determined to have a single Mu d1 insert (Figures 8, 9, and 10) since multiple inserts would confound *lacZ* expression results, and impede the determination of a definitive restriction map for the area of the insert. The maps were compared with the known restriction map of *E. coli*, and several candidate genes were identified by identical

restriction patterns. Most of the identified genes (**Table 6**) are involved in metabolism, but none are regulated by arabinose or glucose (**Table 2** lists some of the expected genes). It is interesting to note that apparent insertions in the *ara* operon were not found.

Considering that restriction mapping possesses inherent imprecision in determining the exact fragment sizes, subcloning and sequencing of the Mu dl:: *E.coli* junction is preferred. Another strategy to confirm the involvement of the cloned gene would be to perform a plasmid replacement assay. By transforming with, and selecting for, a different plasmid (*eg.* pBR322kan) that is incompatible with pBADCm, clones can be rechecked for arabinose or glucose inducibility. This would distinguish between arabinose/glucose affected genes, and those that are affected by Nlp.

Since this series of experiments yielded eight potential clones using a vector without the cloned gene of interest, it can be presumed that libraries should be subject to large-scale screens using selected vectors, so that the selection of non- Nlp- inducible genes may become less significant. It is perhaps not surprising that we identified arabinose- or glucose- inducible genes when using an arabinose-inducible vector. Regardless, none of the genes identified here were known to be induced by either of these sugars.

Mu dl insertions into essential genes would not be isolated due to the lethality of their disruption. To overcome this problem, 2D gels are a viable option (O'Farrell, 1977). By using a reference gel, the induction of numerous genes can be studied simultaneously. Moreover, a knowledge of the regulatory phenotype is not required.

Screening a *lacZ* fusion library, via blue/white selection may fail to identify some genes due to weak beta-galactosidase expression, whereas 2D gels may fail to detect

proteins that are expressed at low levels. Despite these problems, both gene fusion libraries and 2D gels are useful tools for classifying regulatory proteins.

After Nlp-responsive genes have been identified using a *lacZ* transcriptional fusion library, other studies should include the quantification of *lacZ* production (*ie.* the extent of target gene stimulation by Nlp), using liquid β-galactosidase assays. By directly measuring the effect that Nlp has on the expression of this enzyme, then perhaps the effect of Nlp induction (or repression) may be elucidated. Unlike a qualitative assay (Table 4), a quantitative assay will allow for statistical analysis of the amount of target gene induction, or repression, by Nlp.

Moreover, electrophoretic mobility shift assays may be used to understand the exact binding mechanisms of Nlp to the promoters of target genes. It is known that Nlp binds in a manner similar to that of CRP to the CRP binding site (MacIntyre and DuBow, in preparation). These studies may elucidate position dependent, or indirect effects, of Nlp's proposed regulatory properties.

Nlp is believed to function in the regulation of transcription of several diverse genes and regulons. Characterization and identification of Nlp-responsive genes will clarify the role Nlp within wild type *E. coli*. Using a transcriptional fusion library, such as the one constructed in this study, would be a wise choice to perform the task.

References

References

Ahmer, B.M.M., vanReeuwijk, J., Timmer,C.D., Valentine, P.J., and Heffron, F. 1988. *Salmonella typhimurium* encodes an SdiA homolog, a putative quorum sensor of the LuxR family, that regulates genes on the virulence plasmid. Journal of Bacteriology. **180**: 1185-1193

Aiba, H., Nakamura, T., Mitani, H., and Mori, H. 1995. Mutations that alter the allosteric nature of cAMP receptor protein of *Escherichia coli*. The EMBO Journal. **4**:3329-3332

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J.D. 1994. Molecular Biology of the Cell (3rd edition). Garland Publishing, New York, N.Y. pp.401-470

Autexier, C. 1991. Characterization of the Ner/Nlp family of DNA binding regulatory proteins. Microbiology and Immunology. Montreal, McGill University. Doctoral Thesis.

Autexier, C., and DuBow, M.S. 1992. The *Escherichia coli* gene *nlp* is transcribed and evolutionarily conserved among the *Enterobacteriaceae*. Gene. **111**:13-18

Ball, C.A., Osuna, R., Ferguson, K.C., and Johnson, R.C. 1992. Dramatic changes in FIS levels upon nutrient upshift in *Escherichia coli*. Journal of Bacteriology. **174**: 8043-8056

Benevides, J.M., Kukolj, G., Autexier, C., Aubrey, K.L., DuBow, M.S., and Thomas, G.J. 1994. Secondary structure and interaction of phage D108 Ner repressor with a 61-base-pair operator: evidence for altered protein and DNA structure in the complex. Biochemistry. **33**: 10701-10710

Bloch, P.L., Phillips, T.A., and Neidhardt, F.C. 1980. Protein identifications of O'Farrell two-dimensional gels: location of 81 *Escherichia coli* proteins. Journal of Bacteriology. **141:** 1409-1420

Bremer,E., Silhavy, T.J., Weisemann, J.M., and Weinstock, G.M. 1984. λp*lac*Mu: a transposable derivative of bacteriophage lambda for creating *lacZ* protein fusions in a single step. Journal of Bacteriology. **158**: 1084-1093

Bukhari, A.I. and Metlay, M. 1973. Genetic mapping of prophage Mu. Virology. **54**: 109-116

Cai, J. and DuBow, M.S. 1996. Expression of the *Escherichia coli* chromosomal *ars* operon. Canadian Journal of Microbiology. **42**: 662-671

Casadaban, M.J. 1976. Transposition and fusion of the *lac* genes to selected promoters in *Escherichia coli* using bacteriophage lambda and mu. Journal of Molecular Biology. **104**: 541-555

Casadaban, M.T., and Cohen, S.N. 1979 Lactose genes fused to exogenous promoters in one step using a Mu *lac* bacteriophage: *in vivo* probe for transcriptional control sequence. Proceedings of the National Academy of Sciences of the U. S. A. **76**:4530-4533

Castilho, B.A., Olfson, P., and Casadaban, M.J. 1984. Plasmid insertion mutagenesis and *lac* gene fusion with mini-Mu bacteriophage transposons. Journal of Bacteriology. **158**: 488-495

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W., and Prasher, D.C. 1994. Green fluorescent protein as a marker for gene expression. Science. **263**: 802-805

Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, X.C., Stern, D., Winkler, J., Lockhart, D.J., Morris, M.S., and Fodor, S.P.A. 1996. Accessing genetic information with high-density arrays. Science. **274**: 610-614

Choi, Y.L., Nishida, T., Kawamukai, M., Utsumi, R., Sakai, H., and Komano, T. 1989. Cloning and sequencing of an *Escherichia coli* gene, *nlp*, highly homologous to the *ner* genes of bacteriophages Mu and D108. Journal of Bacteriology. **171**:5222-5225

DuBow, M.S. 1994. "Mu and related bacteriophages" from Encyclopedia of Virology. Academic Press, London. Pp.868-876

Engebrecht, J., Simon, M., and Silverman, M. 1985. Measuring gene expression with light. Science. **227**: 1345-1347

Engebrecht, J., Nealson, K., and Silverman, M. 1983. Bacterial bioluminescence: isolation and genetic analysis of functions from *Vibrio fischeri*. Cell. **32**: 773-781

Fodor, S.P.A. 1997. Massively parallel genomics. Science. **277**: 393-395

Fodor, S.P.A. 1993. Multiplexed biochemical assays with biological chips. Nature. 364: 555-556

Fodor, S.P.A., Read, J.L., Pirrung, M.C., Stryer, L., Lu, A.T., and Solas, D. 1991. Light-directed, spatially addressable parallel chemical synthesis. Science. **251**: 767-773

Garcia, J.A., Ou, S.H.I., Wu, F., Lusis, A.J., Sparkes, R.S., and Gaynor, R.B. 1992. Cloning and chromosomal mapping of a human immunodeficiency virus 1 "TATA" element modulatory factor. Proceedings of the National Academy of Sciences of the U.S.A. **89**:9372-9376

Guzman, L.M., Belin, D., Carson, M.J., and Beckwith, J. 1995. Tight regulation, and high-level expression by vectors containing the arabinose P_{BAD} promoter. Journal of Bacteriology. **177**:4121-4130

Guzzo, A., and DuBow, M.S. 1994. A *luxAB* transcriptional fusion to the cryptic *celF* gene of *Escherichia coli* displays increased luminescence in the presence of nickel. Molecular and General Genetics. **242**: 45-460

Hanahan, D. 1983. Studies on transformation of *Escherichia coli* with plasmids. Journal of Molecular Biology. **166**: 557-580

Hastings, J.W. 1996. Chemistries and colours of bioluminescent reactions: a review. Gene. **173**: 5-11

Heim, R., Prasher, D.C., and Tsien, R.Y. 1994. Wavelength mutations and posttranslational autoxidation of green fluorescent protein. Proceedings of the National Academy of Sciences of the United States of America. **91**: 12501-12504

Holloway, B.W. 1993. Genetics for all bacteria. Annual Review of Microbiology. **47**: 659-684

Johnston, M. 1998. Gene chips: array of hope for understanding gene regulation. Current Biology. **8**: R171-R174

Kano, Y, and Imamoto, F. 1990. Requirement of integration host factor (IHF) for growth of *Escherichia coli* deficient in HU protein. Gene. **89:** 133-137

Kano, Y., Yasuzawa, K., Tanaka, H., and Imamoto, F. 1993. Propagation of phage Mu in IHF-deficient *Escherichia coli* in the absence of the HN-S histone-like protein. Gene. **126**: 93-97

Kawamukai, M., Utsumi, R., Takeda, A., Higashi, H., Matsuda, Y.L., Choi, Y.L. and Komano, T. 1991. Nucleotide sequence and characterization of the *sfs1* gene: *sfs1* is involved in CRP*-dependent *mal* gene expression in *E.coli*. Journal of Bacteriology. **173**:2644-2648

Kremer, L., Baulard, A., Estaquier, J., Poulain-Godefroy, and Locht, C. 1995. Green fluorescent protein as a new expression marker in mycobacteria. Molecular Microbiology. **17**: 913-922

Kukolj, G., and DuBow, M.S. 1992. Integration host factor activates the Ner-repressed early promoter of transposable Mu-like phage D108. Journal of Biological Chemistry. **267**:17827-17835

Lemaux, P.G., Herendeen, S.L., Bloch, P.L., and Neidhardt, F.C. 1978. Transient rates of synthesis of individual polypeptides in *E. coli* following temperature shifts. Cell. **13**: 427-434

Lipshutz, R.J., Morris, D., Chee, M., Hubbell, E., Kozal, M.J., Shah, N., Shen, N., Yang, R., and Fodor, S.P.A. 1995. Using oligonucleotide probe arrays to access genetic diversity. Biotechniques. **19**: 442-447

Lockhart, D.J., Dong, H., Byrne, M.C., Follettie, M.T., Gallo, M.V., Chee, M.S., Mittman, M., Wang, C., Kobayashi, M., Horton, H., and Brown, E.L. 1996. Expression monitoring by hybridization to high-density oligonucleotide arrays. Nature Biotchnology. **14**: 1675-1680

Miller, J. 1992. A Short Course in Bacterial Genetics. Cold Spring Harbor Press, New York, NY

Milnes, K. 1994. Honours Research Thesis. McGill University.

Morelle, G. 1989. A plasmid extraction procedure on a mini prep scale. Focus. **11**: 7-8

Neidhardt, F.C. (editor-in-chief). 1996. *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology (2^{nd} edition). ASM Press, Washington D.C. pp.1225-1310; 2588-2612

O'Connor, M.B., and Malamy, M.H. 1983. A new insertion sequence, IS*121*, is found on the Mu d1 (Ap *lac*) bacteriophage and the *Escherichia coli* K-12 chromosome. Journal of Bacteriology. **156**: 669-679

Oehler, S., Eismann, E.R., Kramer, H., and Muller-Hill, B. 1990. The three operators of the *lac* operon cooperate in repression. The EMBO Journal. **9**: 973-979

O'Farrell, P.H. 1975. High resolution two-dimensional electrophoresis of proteins. Journal of Biological Chemistry. *250*: 4007-4021

O'Farrell, P.Z., Goodman, H.M., and O'Farrell, P.H. 1977. High resolution two-dimensional electrophoresis of basic as well as acidic proteins. Cell. **12**: 1133-1142

Pabo, C.O., and Sauer, R.T. 1992. Transcription factors: structural families and principles of DNA recognition. Annual review of Biochemistry. **61**: 1053-1095

Pardee, A.B., Jacob, F., and Monod, J. 1959. The genetic control and cytoplasmic expression of inducibility in the synthesis of β-galactosidase by *E.coli*. Journal of Molecular Biology. **1**: 165

Pease, A.C., Solas, D., Sullivan, E.J., Cronin, M.T., Holmes, C.P., and Fodor, S.P.A. 1994. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. Proceedings of the National Academy of Sciences of the United States of America. **91**: 5022-5026

Prasher, D.C. 1995. Using GFP to see the light. Trends in Genetics. **11**: 320-323

de Saizieu, A., Certa, U., Warrington, J., Gray, C., Keck, W., and Mous, J. 1998. Bacterial transcript imaging by hybridization of total RNA to oligonucleotide arrays. Nature Biotechnology. **16**: 45-48

Sambrook, J., Fritsch, E.F., and Maniatis, T. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor, Cold Spring Harbor Laboratory Press

Sapolsky, R.J., and Lipshutz, R.J. 1996. Mapping genomic library clones using oligonucleotide arrays. Genomics. **33**: 445-456

Shapiro, S.K., Chou, J., Richard, F.V., and Casadaban, M.J. 1983. PMC1871: a protein expression vector. Gene. **25**:71

Shauer, A., Ranes, M., Santamaria, R., Guijarro, J., Lawlor, E., Mendez, C., Chater, K., and Losick, R. 1988. Visualizing gene expression in time and space in the filamentous bacterium *Streptomyces coelicolor*. Science. **240**: 768-772

Shoemaker, D.D., Lashkari, D.A., Morris, D., Mittman, M., and Davis, R.W. 1996. Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy. Nature Genetics: **14**: 450-456

Silhavy, T.J. and Beckwoth, J.R. 1985. Uses of *lac* fusions for the study of biological problems. Microbiological Reviews. **49**: 398-418

Simon, R., Quandt, J., and Klipp, W. 1989. New derivatives of transposon Tn5 suitable for mobilization of replicons, generation of operon fusions, and identification of genes in Gram-negative bacteria. Gene. **80**: 161-169

Sohasky, C.D., Im, H., and Schauer, A.T. 1992a. Construction and application of plasmid- and transposon-based promoter-probe vectors for *Streptomyces* spp. that employ a *Vibrio harveyi* luciferase reporter cassette. Journal of Bacteriology. **174**: 367-376

Sohasky, C.D., Im, H., Nelson, A.D., and Schauer, A.T. 1992b. Tn4556 and luciferase: synergistic tools for visualizing transcription in Streptomyces. Gene. **115**: 67-71

Stearns, T. 1995. The green revolution. Current Biology. **5**: 262-264

Sutcliffe, J.G. 1979. Complete nucleotide sequence of *Escherichia coli* plasmid pBR322. Cold spring Harbor Symposium on Quantitative Biology. **43**:79-90

Ulijasz, A.T., Grenader, A., and Weisblum, B. 1996. A vancomycin-inducible LacZ reporter system in *Bacillus subtilis*: induction by antibiotics that inhibit cell wall sysnthesis by lysozyme. Journal of Bacteriology. **178**: 6305-6309

Viera, J., and Messing, J. 1987. Production of single stranded plasmid DNA. Methods in Enzymology. **153**: 3-11

VonBogelen, R.A., Abshire, K.Z., Putsemlidis, A., Clark, R.L., and Neidhardt, F.C. 1996. P2067-2117. Gene-protein database of *Escherichia coli* K-12,edition 6. In *E. coli* and *Salmonella*: Cellular and Molecular Biology , volume 2. American Society for Microbiology Press, Washington DC.

Wang, S. and Hazelrigg, T. 1994. Implications for *bcd* mRNA localization from spatial distribution of exu protein in *Drosophila* oogenesis. Nature. 369: 400-403

Webb, C.D., Decatur, A., Teleman, A., and Losick, R. 1995. Use of green fluorescent protein for visualization of cell-specific gene expression and subcellular protein localization during sporulation in *Bacillus subtilis.* Journal of Bacteriology. **177**: 5906-5911

Wolk, C.P., Cai, Y., and Panoff, J.M. 1991. Use of a transposon with luciferase as a reporter to identify environmentally responsive genes in a cyanobacterium. Proceedings of the National Academy of Sciences of the United States of America. **88**: 5355-5359