**UCSF**
**UC San Francisco Electronic Theses and Dissertations**

**Title**
Chimeric arrays of complex carbohydrates

**Permalink**
https://escholarship.org/uc/item/55p2s483

**Author**
Johansen, Eric Bennett

**Publication Date**
2006

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

**Copyright 2006**

**By**

**Eric Bennett Johansen**

## ACKNOWLEDGEMENTS

First and foremost I would like to thank Frank Szoka. Your encouragement and continual interest in my success was contagious. In the hardest times, when none of my experiments were working and even I was fed up with this system it was your unwavering support that kept me going. I would not have been able to bring success to this project without the encouragement of Bradford Gibson, whose experience and scientific instincts have always been a source of inspiration. I have had the opportunity of working under the two best human beings this department has to offer, and I am eternally in your debts.

I would to thank Nancy Phillips, whose unwavering support has helped me through many difficult spots. If everyone were allowed to see the world through Nancy's eyes for just one day than the world would be a better place forever.

I would like to thank Mike Apicella, a man of nearly limitless knowledge regarding microbiology. I think you may know more about *Haemophilus influenzae* than it does. You were right all along about the extra KDO.

I would also like to thank Bob Stull, whose skill with Molecular Biology is unparalleled. As a biochemist struggling to impersonate a molecular biologist I benefited greatly from your tutelage, and I thank you. I would also like to thank Mel Sunshine, Josh Park and Tony Zaleski for many helpful conversations along those lines as well.

I want to thank Simon Allen, Rich Cohen, Birgit Schilling, Deb Post, Virginia Platt, Jason Held, Marc Anderson, Gigi Knudsen, Ed Dy, Weijun Li, Zhaohua Hoang, Andrew McKay, Dipali Rhuela, Doug Watson, Grace Huynh, Kareen Reviere, Becky Cebula, and Luise Sternberg for many helpful discussions.

I would like to thank Ted Holman, for encouraging me to go to graduate school and for showing me how rewarding a life in science can be.

I would like to thank Rachael Taylor, my light in the darkness, my partner in crime, my other self. The existence of just one human being as pure of heart as Rachael Taylor is justification for the continuation of the human race. Your endless support and editing skills are as responsible for my success as any of my own contributions.

I want to thank my mom, Cheri Johansen, for raising me to be so sure-footed and strong. For teaching me to stand up and work hard for something I believe in. For insisting that I go to a better undergraduate school than the one I wanted to attend.

I would also like to thank my Father, Bill Johansen. It was his overwhelming insistence that I go to graduate school that ultimately drove my decision to do so. I hope that you know how much I love and respect you. I'd want nothing more in the world than for you to have made it to see me turn in this document. I think I've finally learned the real meaning of the word "Rally."

## COLLABORATIONS

I would like to acknowledge the work carried out by co-workers and collaborators that made this research possible. At the time of this writing, Chapter 3 is about to be submitted as a journal article. Chapter 4 will be formatted as a manuscript pending the outcome of a few continuing experiments.

I would like to thank Luise Sternberg, who was a summer student during the a crucial summer of work. Without your help I would not have been able to accomplish all that cloning.

I would like to thank Tony Zaleski and the rest of our collaborators in the Apicella Lab at the University of Iowa. Tony generated the pGEM-LOS-04 –*lsgA*, pGEM-LOS-04 –*lsgAB*, and pGEM-LOS-04 –*lsgAB* + *siaB* constructs. Although these constructs weren't reported on in this thesis, they will be crucial for the success of Chapter 4 as a publishable article.

I would also like to thank Nancy Phillips, who's preliminary work on the chimeric glycosyltransferase expression system paved the way for me to step in and adapt this into a system for defining glycosyltransferase function.

**Chimeric Arrays of Complex Carbohydrates**

By

Eric Johansen

## ABSTRACT

For the work described in this dissertation, I adapt a chimeric expression system to explore the synthetic properties (specificity and selectivity) of glycosyltransferases.

Despite the abundance of complex carbohydrate structures in nature, their study in therapeutics and biology is severely limited by access to suitable quantities of complex carbohydrates, synthetic or otherwise, with which to initiate study. Automated solid-phase methodologies have revolutionized protein and nucleic acid science, but decades of synthetic research have yet to provide robust methods for automated oligosaccharide synthesis. This hindrance is a consequence of the carbohydrate structure itself: monosaccharides contain several hydroxyl groups of similar reactivity. To synthesize even a simple oligosaccharide, each hydroxyl group must be correctly distinguished from the others in order to obtain the correct product. Assuring the correct regiochemistry in the hydroxyl selection process is tedious, and involves the shuffling of dozens of differential protecting groups, even for the synthesis of a short oligomer.

The increasing complexity of synthetic carbohydrate chemistry stands in stark contrast to what nature has been able to accomplish with simplicity and elegance. Glycosyltransferases, the enzymes responsible for catalyzing the formation of glycosidic bonds in nature, feature exacting stereo- and regioselectivity and catalyze their reactions under mild conditions.

The key obstacle to enzymatic carbohydrate synthesis is the limited body of knowledge regarding the specificity and selectivity of the over 200 identified bacterial glycosyltransferases. Examination of microbial glycosyltransferases in their native systems is complicated by phase variation and degeneracy. *In vitro* determination of glycosyltransferase function would require an expensive array of nucleotide-sugar substrates and nascent-carbohydrate chains with which to test the enzymes.

In this work I successfully designed, implemented, and tested an *in vivo* biosynthetic method for evaluating glycosyltransferase function. By expressing glycosyltransferase genes in an *E. coli* system uncomplicated by degeneracy or phase variation, the sole determinant of function was the availability of an array of nascent-chain substrates for the enzymes to both recognize and glycosylate. In order to generate different nascent-chain substrates, four different methods for shuffling glycosyltransferase genes were examined, and a multiplasmid system was selected as the most stable and robust. Using a number of glycosyltransferase genes from the *lsg* locus of *Haemophilus influenzae* as a test of this method, I determined nascent chain specificitiy and selectivity data for five glycosyltransferases of previously unknown function. Additionally, I have developed six new nascent-chain substrates that can be expressed in *E. coli* in order to determine the synthetic properties of other glycosyltransferases.

## Table of Contents


| | | page |
|---|---|---|
| Acknowledgements | | iii |
| Collaborations | | v |
| Abstract | | vi |
| Table of Contents | | viii |
| List of Tables | | x |
| List of Figures | | xi |
| Chapter 1. | *Introduction* | |
| 1.1 | Carbohydrates in Biology | 1 |
| 1.2 | Synthetic Carbohydrate | 4 |
| 1.3 | Biosynthetic Carbohydrate | 5 |
| 1.4 | Glycosyltransferases | 6 |
| 1.5 | Chimeric Arrays of Complex Carbohydrates | 7 |
| 1.6 | Chimeric LPS Expression and Display in *E. coli* | 8 |
| 1.7 | Mechanism of LPS Expression and Display | 9 |
| 1.8 | LPS Expression/ Display and Chimeric Arrays of Complex Carbohydrates | 11 |
| References | | 12 |
| Chapter 2. | *Cloning Scaffolds for Carbohydrate Synthesis* | |
| Abstract | | 15 |
| Introduction | | 16 |
| Materials and Methods | | 23 |
| Results | | 37 |
| Discussion | | 42 |
| References | | 45 |
| Chapter 3. | *Functional exploration of the Haemophilus influenzae lsg gene locus glycosyltransferases lsgC, lsgD, lsgE, and lsgF, via chimeric expression in Eschericia coli K-12* | |
| Abstract | | 47 |
| Introduction | | 49 |
| Materials and Methods | | 51 |
| Results | | 63 |
| Discussion | | 73 |
| Conclusions | | 76 |
| References | | 79 |

Chapter 4. *Functional determination of the Haemophilus influenzae sialyltransferase lsgB via expression in Escherichia coli K-12*

Abstract 82
Introduction 83
Materials and Methods 87
Results 91
Discussion 95
Conclusions 96
References 97

Chapter 5. *Chimeric conclusions and future experiments*

Conclusions 99
Future Experiments 103
References 105

Appendix

A. Structures of bacterial LPS components. 108
B. Elemental compositions and masses of bacterial LPS components. 109
C. Partially methylated alditol acetates (PMAAs) derived from glycosidically linked sugar residues present in bacterial LPS. 110
D. Haemophilus influenzae lsg locus gene accession numbers. 110
E. Improved protocol for isolation and purification of chimeric E. coli LPS. 111
F. Important concerns for chimeric E. coli LPS preparations. 112
G. Figure A-2: Negative-ion MALDI-TOF spectra from three preparations of O-deacylated LPS from E. coli K-12 transformant LOS-04 113
H. Figure A-2: Negative-ion MALDI-TOF spectra from three preparations of O-deacylated LPS from E. coli K-12 transformant LOS-05 114

**List of Tables**


Chapter 1. page

1-1 A Few Therapeutically Relevant Complex Carbohydrate Sequences 2

Chapter 2.

2-1 Bacterial Strains and Vectors 23
2-2 Primers for Not I Study 24
2-3 Primers for Dra III Study 30
2-4 Primers for the Multiplasmid System 34
2-5 Multiplasmid System *E. coli* Transformants 36

Chapter 3.

3-1 Chimeric oligosaccharides from Phillips et al. 50
3-2 Bacterial Strains and Vectors 52
3-3 Primers used in cloning of the *lsg* genes 53
3-4 Chimeric *E.coli* Transformants 57
3-5 Molecular weights and proposed compositions of the HF treated, O-deacylated LPS from *E. coli* transformants CG, DG, EG, FG, CFG, DFG, EFG, CEFG, DEFG, and LOS-05 65
3-6 ELISA of Mab 3F11 and Mab 6B4 LPS Epitopes on glycoform E4 containing transformants with and without neuraminidase treatment 71
3-7 Methylation analysis of the dephosphorlyated oligosaccharide fractions 73
3-8 Specificity and selectivity conclusions 76

Chapter 4.

4-1 Chimeric oligosaccharides from previous chimeric studies 83
4-2 Bacterial strains and vectors 87
4-3 Chimeric *E. coli* transformants 88
4-4 Molecular weights and proposed compositions of the O-deacylated LPS from *E. coli* transformants LOS-05, LOS-04, and LOS-04 + *SiaB* 93

Chapter 5.

5-1 Specificity and Selectivity Conclusions 102

Appendix.

A1 Structures of bacterial LPS components 108
A2 Partially methylated alditol acetates (PMAAs) derived from glycosidically linked sugar residues present in bacterial LPS 110
A3 *Haemophilus influenzae lsg* locus gene accession numbers 110

**List of Figures**


Chapter 1 page

1-1 Proposed pathway for the synthesis of chimeric LPS and *E. coli* K-12 10

Chapter 2

2-1 Graphical depiction of combinatorial Not 1 ligation 18
2-2 Graphical depiction of phosphorylation - limited 3-gene ligation 19
2-3 The Dra III restriction site used in a poly gene single step ligation 21
2-4A Schematic drawing of PGEM-LOS-10 25
2-4B Schematic drawing of PGEM-LOS-12 33
2-5 Agarose gel of plasmid DNA from combinatorial ligation 38
2-6 Agarose gel of gene DNA from phosphorylation-limited ligation 40

Chapter 3

3-1 Negative ion MALDI-TOF spectra of the O-deacylated, HF treated LPS from transformants CG, DG, EG, and FG. 64
3-2 Negative ion MALDI-TOF spectra of the O-deacylated, HF treated LPS from transformants CFG, DFG, and EFG. 68
3-3 Negative ion MALDI-TOF spectra of the O-deacylated, HF treated LPS from transformants CEFG, DEFG, and LOS-05. 70

Chapter 4

4-1 Structure of NeuAc and both conformational representations of KDO 85
4-2 Negative ion MALDI-TOF spectra of the O-deacylated LPS from *E. coli* K-12 transformants LOS-05, LOS-04, and LOS-04 + *SiaB* 92

Chapter 5

5-1A/B Chimeric LPS structures synthesized as part of this work 100

Appendix

A-1 Structures of bacterial LOS components 108
A-2 Negative-ion MALDI-TOF spectra from three preparations of O-deacylated LPS from *E. coli* K-12 transformant LOS-04 113
A-3 Negative-ion MALDI-TOF spectra from three preparations of O-deacylated LPS from *E. coli* K-12 transformant LOS-04 114

# CHAPTER 1.

# INTRODUCTION

## 1.1 Carbohydrates in Biology

Carbohydrates are ubiquitous in biology. In eukaryotes, they function as metabolites, structural elements, and signaling molecules and are involved in inflammation, immunological responses, and metastases, among others. Furthermore, carbohydrates are medicinally important biomolecules [1-10].

Conjugated to proteins to form glycoproteins, complex carbohydrates alter protein structure and function. Among known methods of post-translational modification, glycosylation is the most ubiquitous and complex. Greater than half of known human proteins are glycosylated [2,9,11-14]. Conjugated to lipids to form glycolipids, saccharides have pivotal roles in cell-cell recognition and signaling [2,9]. The extracellular matrix also contains large molecules called proteoglycans that play a part in providing biomechanical structure and strength to organs and tissues. They also have the capability to participate both directly and indirectly in cell signaling processes [9,15].

In addition to their many functions in eukaryotes, complex carbohydrates are surface components of many microbial pathogens. These pathogens express complex carbohydrate on their surfaces, mimicking the saccharide-rich mucin layers that coat epithelial cells, and/or various glycoconjugates present on nearly every host cell surface [7,16,17]. A number of bacteria that colonize human mucosal surfaces exclusively express carbohydrate antigens that are molecular mimics of glycosphingolipids found on human cells. Pathogenic strains of *Neisseria gonorrhoeae, Neisseria meningitides* and

*Haemophilus influenzae* decorate their surfaces with molecules termed lipooligosaccharides, or LOS. These LOS are antigenically and chemically similar to lactoneoseries glycosphingolipids [7,17].

Carbohydrate structures are important markers for other biological processes and disease states as well. See Table 1-1 for a summary of just a few therapeutically relevant glycoforms.

## 1.2 Synthetic Carbohydrate

Despite the abundance of complex carbohydrate structures, their study in therapeutics and biology is severely limited by access to suitable quantities of complex carbohydrates, synthetic or otherwise. Obtaining complex carbohydrates is much more complicated than procuring peptides/proteins and oligonucleotides/DNA. Automated solid-phase methodologies have revolutionized protein and nucleic acid science, but decades of synthetic research have yet to provide robust methods for automated oligosaccharide synthesis. This hindrance is a consequence of the carbohydrate structure itself: monosaccharides contain several hydroxyl groups of similar reactivity. To synthesize even a simple oligosaccharide, each hydroxyl group must be correctly distinguished from the others in order to obtain the correct product. Assuring the correct regiochemistry in the hydroxyl selection process is tedious, and involves the shuffling of dozens of differential protecting groups, even for the synthesis of a short oligomer. Furthermore, controlling the anomeric stereochemistry of the product is also very

complicated, as the formation of glycosidic bonds can go through either an $S_N1$ or $S_N2$ reaction.

This is not to say that carbohydrate synthesis is impossible. Danishefsky and coworkers have prepared several novel, fully synthetic, carbohydrate-based antitumor vaccines [8,18]. Wong and co-workers have developed a novel carbohydrate synthetic method. They use a computer program called Optimer that stores the reactivity profiles of a number of anomerically activated sugar molecules that are protected at all but one hydroxyl. In a one-pot reaction scheme the monomers are added in order of decreasing reactivity [6,9,11]. There is no need for selective de-protection between steps, however you must have a pool of selectively de-protected monomers to begin with.

Despite these and other advances, carbohydrate synthesis remains in the hands of a small group of elite synthetic laboratories, unlike the synthesis of oligonucleotides and peptides, which can be obtained readily and inexpensively by virtually all scientists.

### 1.3 Biosynthetic Carbohydrate

The increasing complexity of synthetic carbohydrate chemistry stands in stark contrast to what nature has been able to accomplish with simplicity and elegance. The past few decades have given rise to a number of enzymatic syntheses strategies for the production of complex carbohydrate. Glycosyltransferases, the enzymes responsible for catalyzing the formation of glycosidic bonds in nature, feature exacting stereo- and regioselectivity and catalyze their reactions under mild conditions [6,11]. There is no need for protecting groups, and the enzyme orchestrates the stereochemistry of the

anomeric configuration. However, the activated nucleotide-sugars that are the substrates for these enzymes are often unstable and expensive.

A technique that elegantly side-steps this particular difficulty, bacterial metabolic pathway engineering, has recently emerged as a powerful method for the large-scale synthesis of oligosaccharides [12,27,28]. Whole cells, rather than individual enzymes perform glycosylation reactions. Microbial glycosyltransferases of interest are cloned into a vector and expressed in *E. coli* just as one might overexpress a protein. The nucleotide-sugar substrates for the glycosyltransferases are quite often produced by the metabolism of the host bacterium. In the cases where they are not available, the relevant sugar-nucleotide synthetase can usually be cloned alongside the glycosyltransferase of interest [12].

### 1.4 Glycosyltransferases

The key limitation to enzymatic and bacterial metabolic pathway engineering carbohydrate synthesis is the limited body of knowledge regarding the specificity and selectivity of the over 200 identified bacterial glycosyltransferases [29].

The difficulty in evaluating their substrate specificity is due to the nature of the glycosyltransferases themselves. Glycosyltransferase genes typically operate with three-fold specificity. They are specific to the type of monomer they add to a nascent chain of growing oligosaccharide, they are specific with respect to the type of linkage they form between this monomer and the nascent chain, and they are specific to a number of residues on the growing nascent chain (typically a disaccharide) [29,30].

Examination of the function of microbial glycosyltransferases in their endogenous systems is complicated by degeneracy and phase variation, a system with which a bacterium can alter its glycosylation patterns to adapt to its environment [31-34].

In order to evaluate the functions of microbial glycosyltransferases *in vitro*, first the gene would have to be cloned, expressed, and purified from the rest of the contents of the cytoplasm. Characterization of the resulting purified glycosyltransferase would then require the acquisition of a range of nucleotide-sugar substrates as well as a number of mono-, di- and tri-saccharides of varying composition in which to test the enzyme's specificities.

## 1.5 Chimeric Arrays of Complex Carbohydrates

A bacterial metabolic pathway engineering approach, in which an array of *E. coli* colonies expressing short carbohydrate chain backgrounds, were co-transfected with glycosyltransferase genes of unknown function, can be used to elucidate the specificities and relationships between glycosyltransferases and their substrates. *E. coli* carbohydrate metabolism will supply most of the relevant nucleotide-sugar substrates for microbial glycosyltransferases. Providing the glycosyltransferases with an adequate selection of nascent-chain carbohydrate substrates would, in turn, yield a system in which each carbohydrate structure elongated would confirm a putative glycosyltransferase function. Furthermore, each system in which the nascent-chain carbohydrate substrates were not elongated would also yield specificity and selectivity data about the glycosyltransferase.

Needed for this approach, however, would be the ability to shuffle glycosyltransferase genes in a way that could lead to the synthesis of defined one and two-sugar nascent oligosaccharide chains. These chains would then provide the means with which to evaluate the nascent chain specificity of further glycosyltransferases expressed in the system and so on.

As the library of *E. coli* colonies expressing short carbohydrate chain backgrounds expanded, it would also benefit from high-throughput screening technology. Small libraries could be screened with conventional methods of extraction, purification, and mass spectrometric analysis. However, a system with a built-in capacity for expansion would be more desirable.

### 1.6 Chimeric LPS Expression and Display in *E. coli*

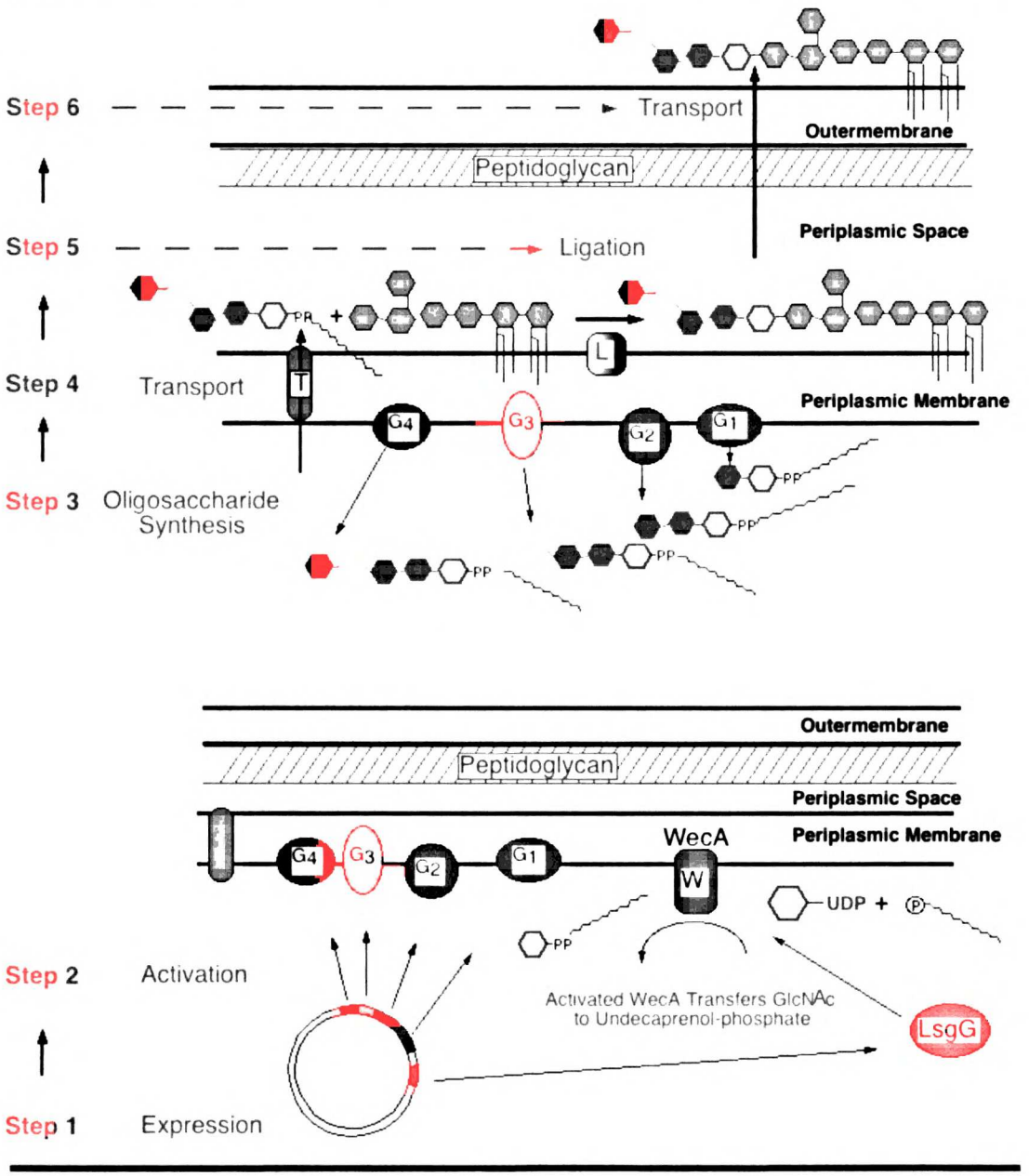
A system that could be adapted to the high-throughput functional exploration of glycosyltransferase genes was first discovered by Apicella and colleagues [35]. In an attempt to clone glycosyltransferase genes into *E. coli* for analysis, they discovered a novel system of carbohydrate expression and display. Essentially, expression of a chimeric polysaccharide in a K-12 strain of *E. coli* was achieved by transformation with exogenous plasmids containing glycosyltransferase and related regulatory genes from the *lsg* (lipopolysaccharide synthesis gene) locus of *Haemophilius influenzae* [35-38]. Data from work in this system by Phillips and colleagues [36] showed that a set of chimeric LPS structures can be produced that extend beyond the rough-LPS (rLPS) core structure produced by the background *E. coli* K-12 strain by the addition of two to five new sugars.

In each case, the precise sugar extension was dependent on which subset of glycosyltransferase genes were present in the plasmid. In all cases, the extended sugar sequences were found to be attached to the terminal non-reducing heptose of the endogenous *E. coli* rough lipopolysaccharide (rLPS) molecule. The protein responsible for initiating synthesis of the oligosaccharide encoded by the exogenous *H. influenzae* genes was identified as WecA, an endogenous *E. coli* protein. Interestingly, this endogenous *E. coli* protein seems to be regulated by the *H. influenzae lsgG* gene product contained in all the functional *H. influenzae* plasmids.

### 1.7 Mechanism of LPS Expression and Display

The putative mechanism of this phenomenon is illustrated in Figure 1-1. The *H. influenzae* LsgG activates the endogenous *wecA* gene. WecA then transfers the first N-acetyl-glucosamine from its activated uridine diphosphate carrier to an undecaprenol-phosphate molecule. The fatty tail of the undecaprenol associates the nascent lipid-oligosaccharide to the cytoplasmic face of the inner membrane, which is also the site of the exogenous glycosyl transferase gene products. The glycosyltransferases then add their respective sugar monomers to the nascent chain. After the chain is synthesized it is exported into the periplasmic space, where it is ligated to the endogenous *E. coli* rLPS (synthesized through an alternate pathway entirely). After the ligation step it is transported through the peptidoglycan layer and comes to reside as a main constituent of the extracellular face of the outer membrane [31,36,39-41].

***Figure 1-1: Proposed pathway for the synthesis of chimeric LPS in E. coli K-12***



## 1.8 LPS Expression/Display and Chimeric Arrays of Complex Carbohydrates

This LsgG - *E. coli* LPS expression/display system fulfills three of the key requirements for a system needed to explore the function of microbial glycosyltransferase genes for synthetic purposes. 1.) *E. coli* metabolism supplies most of the nucleotide-sugar substrates for the glycosyltransferase genes. 2.) Carbohydrate synthesized in this system is placed onto a convenient scaffold in which the expression of short carbohydrate chain backgrounds can be controlled by the addition of exogenous glycosyltransferase genes. 3.) The exterior display of the synthesized LPS will ultimately allow for expansion of the system into larger libraries. An unknown glycosyltransferase gene cloned into a library of *E. coli* colonies expressing nascent chain backgrounds can quickly be scanned for alterations of carbohydrate phenotype with antibody and/or lectin binding assays.

What I have detailed in this dissertation is the development of a system to shuffle glycosyltransferase genes in a way that leads to the synthesis of meaningful one- and two-sugar nascent oligosaccharide chains in *E. coli* K-12. Furthermore, I describe the use of this system to determine the functional properties of specificity and selectivity for five glycosyltransferase genes from the *lsg* locus of *H. influenzae*. The synthetic functions of the glycosyltransferases explored in this study have been a matter of debate for over fifteen years [7,16,31,35-38,41-45]. The results of these experiments further our understanding of glycosyltransferase functions that would be difficult to elucidate *in vitro* or in a more complicated system.

## REFERENCES


1. Albermann C, Piepersberg W, Wehmeier UF: **Synthesis of the milk oligosaccharide 2'-fucosyllactose using recombinant bacterial enzymes.** *Carbohydr Res* 2001, **334**:97-103.
2. Alper J: **Searching for medicine's sweet spot.** *Science* 2001, **291**:2338-2343.
3. Alper J: **Glycobiology. Turning sweet on cancer.** *Science* 2003, **301**:159-160.
4. Bettler E, Samain E, Chazalet V, Bosso C, Heyraud A, Joziasse DH, Wakarchuk WW, Imberty A, Geremia AR: **The living factory: in vivo production of N-acetyllactosamine containing carbohydrates in E. coli.** *Glycoconj J* 1999, **16**:205-212.
5. Harvey HA, Porat N, Campbell CA, Jennings M, Gibson BW, Phillips NJ, Apicella MA, Blake MS: **Gonococcal lipooligosaccharide is a ligand for the asialoglycoprotein receptor on human sperm.** *Mol Microbiol* 2000, **36**:1059-1070.
6. Koeller KM, Wong CH: **Complex carbohydrate synthesis tools for glycobiologists: enzyme-based approach and programmable one-pot strategies.** *Glycobiology* 2000, **10**:1157-1169.
7. Moran AP, Prendergast MM, Appelmelk BJ: **Molecular mimicry of host structures by bacterial lipopolysaccharides and its contribution to disease.** *FEMS Immunol Med Microbiol* 1996, **16**:105-115.
8. Ragupathi G, Koide F, Livingston PO, Cho YS, Endo A, Wan Q, Spassova MK, Keding SJ, Allen J, Ouerfelli O, et al.: **Preparation and evaluation of unimolecular pentavalent and hexavalent antigenic constructs targeting prostate and breast cancer: a synthetic route to anticancer vaccine candidates.** *J Am Chem Soc* 2006, **128**:2715-2725.
9. Sears P, Wong CH: **Toward automated synthesis of oligosaccharides and glycoproteins.** *Science* 2001, **291**:2344-2350.
10. Zaleski A, Scheffler NK, Densen P, Lee FK, Campagnari AA, Gibson BW, Apicella MA: **Lipooligosaccharide P(k) (Galalpha1-4Galbeta1-4Glc) epitope of moraxella catarrhalis is a factor in resistance to bactericidal activity mediated by normal human serum.** *Infect Immun* 2000, **68**:5261-5268.
11. Hanson S, Best M, Bryan MC, Wong CH: **Chemoenzymatic synthesis of oligosaccharides and glycoproteins.** *Trends Biochem Sci* 2004, **29**:656-663.
12. Antoine T, Priem B, Heyraud A, Greffe L, Gilbert M, Wakarchuk WW, Lam JS, Samain E: **Large-scale in vivo synthesis of the carbohydrate moieties of gangliosides GM1 and GM2 by metabolically engineered Escherichia coli.** *Chembiochem* 2003, **4**:406-412.
13. Prakobphol A, Thomsson KA, Hansson GC, Rosen SD, Singer MS, Phillips NJ, Medzihradszky KF, Burlingame AL, Leffler H, Fisher SJ: **Human low-molecular-weight salivary mucin expresses the sialyl lewisx determinant and has L-selectin ligand activity.** *Biochemistry* 1998, **37**:4916-4927.
14. Sears P, Wong CH: **Intervention of carbohydrate recognition by proteins and nucleic acids.** *Proc Natl Acad Sci U S A* 1996, **93**:12086-12093.
15. Kresse H, Schonherr E: **Proteoglycans of the extracellular matrix and growth control.** *J Cell Physiol* 2001, **189**:266-274.

16. Swords WE, Jones PA, Apicella MA: **The lipo-oligosaccharides of Haemophilus influenzae: an interesting array of characters.** *Journal of Endotoxin Research* 2003, **9**:131-144.
17. Harvey HA, Swords WE, Apicella MA: **The mimicry of human glycolipids and glycosphingolipids by the lipooligosaccharides of pathogenic neisseria and haemophilus.** *J Autoimmun* 2001, **16**:257-262.
18. Danishefsky SJ, Allen JR: **From the Laboratory to the Clinic: A Retrospective on Fully Synthetic Carbohydrate-Based Anticancer Vaccines Frequently used abbreviations are listed in the appendix.** *Angew Chem Int Ed Engl* 2000, **39**:836-863.
19. Prakobphol A, Tangemann K, Rosen SD, Hoover CI, Leffler H, Fisher SJ: **Separate oligosaccharide determinants mediate interactions of the low-molecular-weight salivary mucin with neutrophils and bacteria.** *Biochemistry* 1999, **38**:6817-6825.
20. Kannagi R: **Regulatory roles of carbohydrate ligands for selectins in the homing of lymphocytes.** *Current Opinion in Structural Biology* 2002, **12**:599-608.
21. Swords WE, Buscher BA, Ver Steeg Ii K, Preston A, Nichols WA, Weiser JN, Gibson BW, Apicella MA: **Non-typeable Haemophilus influenzae adhere to and invade human bronchial epithelial cells via an interaction of lipooligosaccharide with the PAF receptor.** *Mol Microbiol* 2000, **37**:13-27.
22. Love DC, Hanover JA: **The hexosamine signaling pathway: deciphering the "O-GlcNAc code".** *Sci STKE* 2005, **2005**:re13.
23. Hanover JA: **Glycan-dependent signaling: O-linked N-acetylglucosamine.** *Faseb J* 2001, **15**:1865-1876.
24. Harvey HA, Jennings MP, Campbell CA, Williams R, Apicella MA: **Receptor-mediated endocytosis of Neisseria gonorrhoeae into primary human urethral epithelial cells: the role of the asialoglycoprotein receptor.** *Mol Microbiol* 2001, **42**:659-672.
25. Watkins WM: **The ABO blood group system: historical background.** *Transfus Med* 2001, **11**:243-265.
26. Eliaz RE, Szoka FC, Jr.: **Liposome-encapsulated doxorubicin targeted to CD44: a strategy to kill CD44-overexpressing tumor cells.** *Cancer Res* 2001, **61**:2592-2601.
27. Fort S, Birikaki L, Dubois MP, Antoine T, Samain E, Driguez H: **Biosynthesis of conjugatable saccharidic moieties of GM(2) and GM(3) gangliosides by engineered E-coli.** *Chemical Communications* 2005:2558-2560.
28. Endo T, Koizumi S: **Large-scale production of oligosaccharides using engineered bacteria.** *Curr Opin Struct Biol* 2000, **10**:536-541.
29. Coutinho PM, Deleury E, Davies GJ, Henrissat B: **An evolving hierarchical family classification for glycosyltransferases.** *J Mol Biol* 2003, **328**:307-317.
30. Breton C, Mucha J, Jeanneau C: **Structural and functional features of glycosyltransferases.** *Biochimie* 2001, **83**:713-718.
31. Preston A, Mandrell RE, Gibson BW, Apicella MA: **The lipooligosaccharides of pathogenic gram-negative bacteria.** *Crit Rev Microbiol* 1996, **22**:139-180.
32. Varki A: **Biological roles of oligosaccharides: all of the theories are correct.** *Glycobiology* 1993, **3**:97-130.

33. Zwahlen A, Rubin LG, Moxon ER: **Contribution of lipopolysaccharide to pathogenicity of Haemophilus influenzae: comparative virulence of genetically-related strains in rats.** *Microb Pathog* 1986, **1**:465-473.
34. Kimura A, Hansen EJ: **Antigenic and phenotypic variations of Haemophilus influenzae type b lipopolysaccharide and their relationship to virulence.** *Infect Immun* 1986, **51**:69-79.
35. Spinola SM, Kwaik YA, Lesse AJ, Campagnari AA, Apicella MA: **Cloning and expression in Escherichia coli of a Haemophilus influenzae type b lipooligosaccharide synthesis gene(s) that encodes a 2-keto-3-deoxyoctulosonic acid epitope.** *Infect Immun* 1990, **58**:1558-1564.
36. Phillips NJ, Miller TJ, Engstrom JJ, Melaugh W, McLaughlin R, Apicella MA, Gibson BW: **Characterization of chimeric lipopolysaccharides from Escherichia coli strain JM109 transformed with lipooligosaccharide synthesis genes (lsg) from Haemophilus influenzae.** *J Biol Chem* 2000, **275**:4747-4758.
37. Phillips NJ, Apicella MA, Griffiss JM, Gibson BW: **Structural studies of the lipooligosaccharides from Haemophilus influenzae type b strain A2.** *Biochemistry* 1993, **32**:2003-2012.
38. Abu Kwaik Y, McLaughlin RE, Apicella MA, Spinola SM: **Analysis of Haemophilus influenzae type b lipooligosaccharide-synthesis genes that assemble or expose a 2-keto-3-deoxyoctulosonic acid epitope.** *Mol Microbiol* 1991, **5**:2475-2480.
39. Whitfield C, Paiment A: **Biosynthesis and assembly of Group 1 capsular polysaccharides in Escherichia coli and related extracellular polysaccharides in other bacteria.** *Carbohydr Res* 2003, **338**:2491-2502.
40. Raetz CR, Whitfield C: **Lipopolysaccharide endotoxins.** *Annu Rev Biochem* 2002, **71**:635-700.
41. Phillips NJ, John CM, Reinders LG, Gibson BW, Apicella MA, Griffiss JM: **Structural models for the cell surface lipooligosaccharides of Neisseria gonorrhoeae and Haemophilus influenzae.** *Biomed Environ Mass Spectrom* 1990, **19**:731-745.
42. Jones PA, Samuels NM, Phillips NJ, Munson RS, Jr., Bozue JA, Arseneau JA, Nichols WA, Zaleski A, Gibson BW, Apicella MA: **Haemophilus influenzae type b strain A2 has multiple sialyltransferases involved in lipooligosaccharide sialylation.** *J Biol Chem* 2002, **277**:14598-14611.
43. Hood DW, Randle G, Cox AD, Makepeace K, Li J, Schweda EKH, Richards JC, Moxon ER: **Biosynthesis of Cryptic Lipopolysaccharide Glycoforms in Haemophilus influenzae Involves a Mechanism Similar to That Required for O-Antigen Synthesis.** *J. Bacteriol.* 2004, **186**:7429-7439.
44. Phillips NJ, McLaughlin R, Miller TJ, Apicella MA, Gibson BW: **Characterization of two transposon mutants from Haemophilus influenzae type b with altered lipooligosaccharide biosynthesis.** *Biochemistry* 1996, **35**:5937-5947.
45. Mandrell RE, McLaughlin R, Aba Kwaik Y, Lesse A, Yamasaki R, Gibson B, Spinola SM, Apicella MA: **Lipooligosaccharides (LOS) of some Haemophilus species mimic human glycosphingolipids, and some LOS are sialylated.** *Infect Immun* 1992, **60**:1322-1328.

# CHAPTER 2.

# CLONING SCAFFOLDS FOR CARBOHYDRATE SYNTHESIS

## ABSTRACT

Seven genes in the *Haemophilus influenzae lsg* gene locus have been shown to express extended lipopolysaccharide (LPS) structures in *E. coli* when transfected *en masse* or in small groups [1]. However, the functions of the individual genes had yet to be elucidated. To determine the functions of these genes, I have explored four different cloning strategies for the purposes of shuffling entire open reading frames in an *Eschericia coli* complex carbohydrate expression system. They are as follows: A.) "Combinatorial ligation" of an equimolar mixture of genes into a compatible vector at different gene to vector ratios; B.) A novel method:"Phosphorylation-Limited 3-Gene Ligation;" C.) A "Hapaxomer" method, using the Hapaxoterministic Endonuclease Dra III and designer sticky ends [2,3]; and D.) A "Multiplasmid" method using traditional molecular biology techniques to clone individual genes into vectors, achieving combination by transformation of the *E. coli* with multiple vectors. Positive and negative aspects of each system for the purpose of expressing a set of glycosyltransferase genes in concert will be discussed. The "Multiplasmid" method, which combines the genes at the level of transformation rather than ligation, was the most robust and reliable method explored. It was sufficiently versatile and allowed for successful phenotype analysis of a set of varied glycosyltransferase gene combinations.

## INTRODUCTION

Examination of the function of microbial glycosyltransferase genes in their endogenous systems is complicated by degeneracy and phase variation, a system with which a bacterium can alter its glycosylation patterns to adapt to its environment [4-7]. Our lab had pioneered an expression system in which exogenous glycosyltransferases could be expressed in *E. coli* without such complications [1]. Transformation of *E. coli* with the *Haemophilus influenzae lsg* gene locus, or various truncations of the *lsg locus,* alters the phenotype of the *E. coli* lipopolysaccharide (LPS) in a manner proportional to the amount of the locus genes transfected. However, a more versatile system of gene combination was required to delineate the synthetic functions of each of the seven genes from this locus.

The difficulty in evaluating the substrate specificity is due to the nature of the glycosyltransferases themselves. LPS glycosyltransferase genes typically operate with three-fold specificity. They are specific to the type of monomer they add to a nascent chain of growing oligosaccharide, they are specific with respect to the type of linkage they form between this monomer and the nascent chain, and they are specific to a number of residues on the growing nascent chain [8,9]. *E. coli* carbohydrate metabolism supplies the relevant nucleotide-sugar substrates for the glycosyltransferases in this study [10], and the glycosyltransferases themselves confer the linkage specificity. What was needed was the ability to shuffle glycosyltransferase genes in a way that could lead to the synthesis of defined one- and two-sugar nascent oligosaccharide chains. These chains

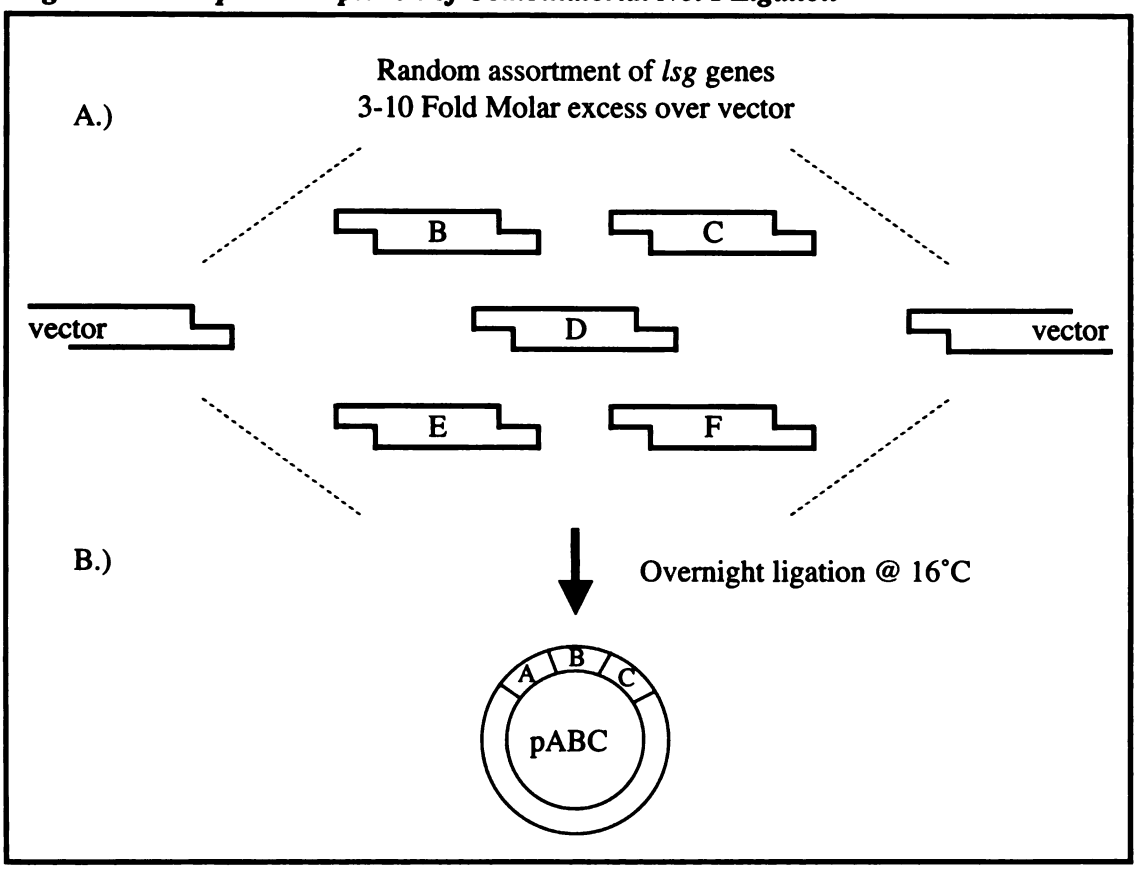
would then provide the means with which to evaluate the nascent chain specificity of additional glycosyltransferases expressed in the system and so on.

Combinatorial techniques such as phage display were becoming popular for modulating small regions of genes and proteins. Santi and co-workers had began investigating domain-swapping methods for shuffling small fragments of polyketide synthesis genes [11]. They had also developed a system that employed random digestion and re-ligation of the genes in an error-prone fashion leading to "mutagenesis" of the domains or modules of a mixture of PKS genes [12]. However, no literature precedent was obvious at the time, for a system of shuffling entire genes in a modular system.

The first method explored for this purpose was the "Combinatorial ligation" method. Primers were designed adding a rare 8bp restriction site (Not I) to the end of the annealing sequences for the PCR amplification of each of five *lsg* genes. A vector was constructed that contained the Not I site and the *lsgG* ORF, which is necessary for the expression of exogenous glycosyltransferase machinery in *E. coli* [1]. PCR of each *lsg* gene, *B-F*, was performed independently and the genes were digested with Not I, quantitated and then mixed together at equimolar concentrations. This mixture was ligated into the Not I site of the vector at gene:vector molar ratios high enough to ensure that more than one gene was ligated into each vector (Figure 2-1). These vectors were directly transfected into *E. coli* and screened for altered phenotypes.

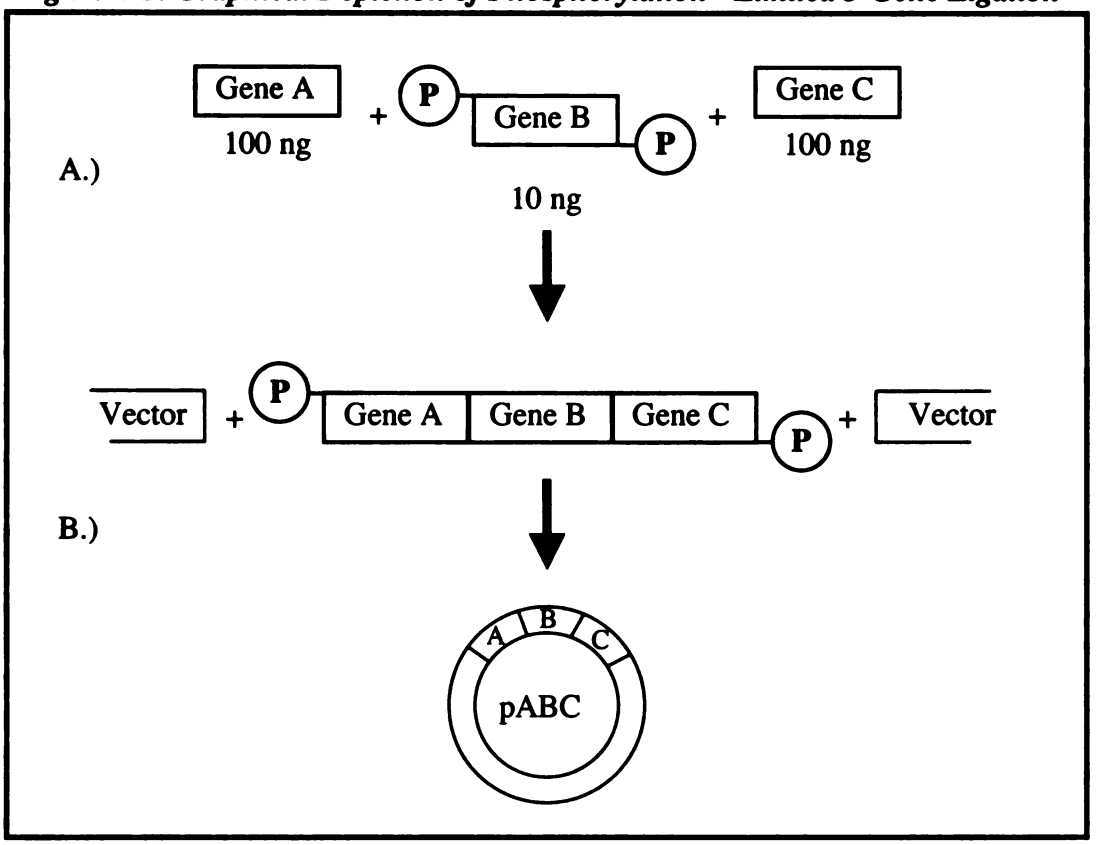
The second method employed in the pursuit of multiple-gene ligation was dubbed "phosphorylation-limited 3-gene ligation" (Figure 2-2). This technique aimed to increase the probability of obtaining multiple-gene ligation products. This was a two-step ligation

***Figure 2-1: Graphical Depiction of Combinatorial Not I Ligation***




A.) Graphical depiction of a mixture of five genes at equimolar concentrations. The genes are mixed in at a 3-10 fold molar excess with cloning vector. All genes and the vector have compatible sticky ends. B.) An overnight ligation should yield vectors with a gene:vector ratio mirroring that of the reaction gene:vector concentration ratio.

***Figure 2-2: Graphical Depiction of Phosphorylation - Limited 3-Gene Ligation***

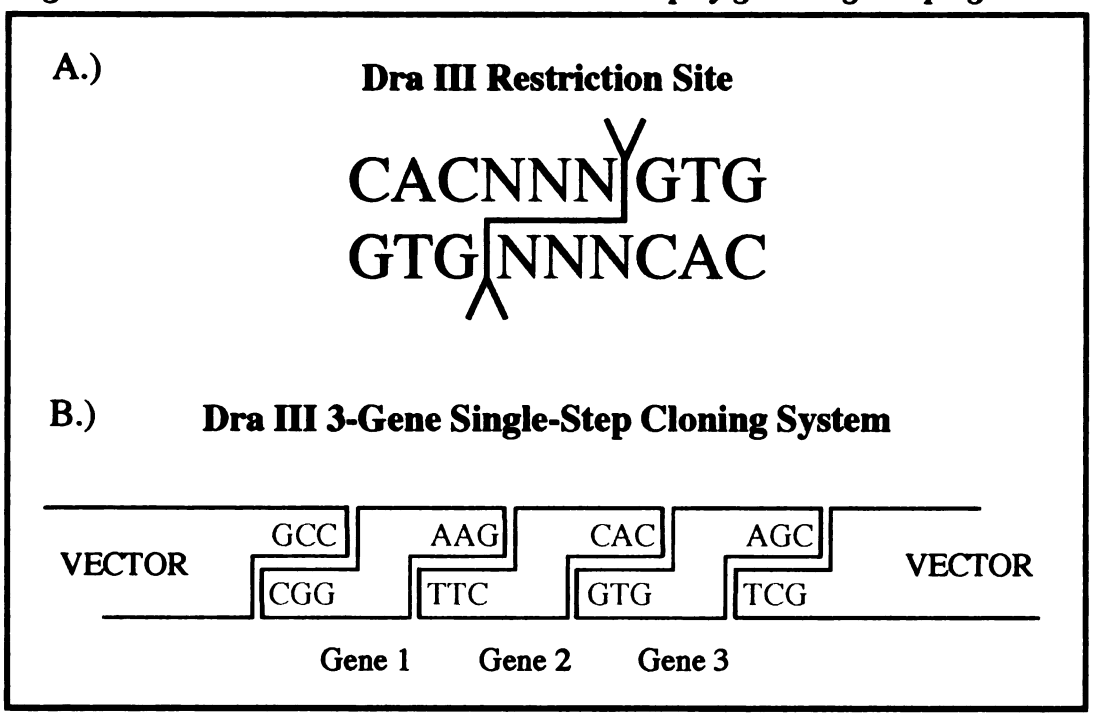



A.) The first step of phosphorylation-limited 3-gene ligation. A limiting concentration of a single phosphorylated gene in a ligation reaction with two dephosphorylated genes forms a 3-gene product. B.) The 3-gene product is phosphorylated and ligated into the vector.

reaction, in which three genes would be ligated together, then separated and visualized via agarose gel electrophoresis. The band containing 3 genes would be excised from the gel, purified, and ligated into the cloning vector. To accomplish this, two of the three genes in the first step of the ligation reaction would be dephosphorylated with calf intestinal phosphatase (CIP) before the ligation. A large molar excess of the dephosphorylated genes ensured that the phosphorylated gene didn't dominate the gene ligation step. Using the same background vector and primers from the first set of experiments, we did a thorough evaluation of this technique for the purposes of cloning 1-3 genes into an expression vector.

The third method described in this study, the "Hapaxomer" method, is a technique first performed by Leibler and coworkers [2]. They employed a combinatorial genetic technique using a new class of restriction endonucleases called "Hapaxoterministic" endonucleases (hapaxomers). Hapaxomers cleave DNA outside the restriction site, or within an interrupted palindrome, at bases that are not specified [3]. The use of a PCR based system and an interrupted palindrome hapaxomer, such as Dra III, would enable us to use "designer sticky ends" to assign the position of a gene in a three-gene cloning system with the sequence of its restriction site (Figure 2-3). Each gene in this system as well as both ends of the linearized vector would only be able to anneal with the fragment of DNA it was intended to anneal with. This would be a one-step ligation without the need to dephosphorylate the vector with CIP, which was suspected of degrading sticky ends in previous ligation attempts. A new vector was constructed with two Dra III sites and an *lsgG* ORF. New primers were ordered for a test set of three genes, and this technique was thoroughly examined.

*Figure 2-3: The Dra III restriction site used in a poly gene single step ligation*




A.) Depiction of the restriction site of the hapaxoterministic endonuclease Dra III. The enzyme recognizes the flanking CAC and GTG sequences but ignores the three base pairs in the cut site. B.) A system of four designer sticky-ends and their complements that could be used to perform a single-step poly-gene ligation.

The fourth and final method investigated was the "Multiplasmid" method. Essentially, traditional molecular biology techniques were used to create single or double-gene vectors with different antibiotic resistances. Transfecting different two-vector combinations into each clone or "transformant" achieved combination. This technique was developed due to a growing suspicion that DNA ligation reactions were insufficiently robust for multiple-gene ligations. Additionally, a study published by Santi and coworkers suggested that combining the functional domains of polyketide synthases at the level of transfection (multiple varied vectors per organism) would be more robust than PCR and ligase-based methods [13]. A system was designed in which each individual glycosyltransferase gene from the *lsg* gene locus would be cloned into the *lsgG* ORF containing vector, pGEM-LOS-12. One and two glycosyltransferase gene vectors were created in another vector, pACYC184, which has an origin of replication that is compatible with the Col E1 origin in the pGEM vector. These vectors were constructed using traditional cloning techniques: one gene was cloned into one vector at a time. The vectors were constructed and evaluated for their stability, ability to alter the phenotype of the *E. coli* LPS, and for their ease of use and combination. All successful single and multi vector transformants are presented in Table 2-5 of the materials and methods section.

## MATERIALS AND METHODS

***Table 2-1 Bacterial Strains and Vectors***

| Strain/Plasmid | Relevant Characteristics | Ref./Source |
|---|---|---|
| E. Coli | | |
| XL-1 Bluescript | *recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F'*proAB lacI*$^{q}$*ZDM15 Tn10* (Tet$^{r}$)] | Stratagene |
| Plasmids | | |
| pGEM3Z f+ | Ap$^{r}$ | Promega Biotech |
| pGEM3Z | Ap$^{r}$ | Promega Biotech |
| pGEM-LOS-4 | Ap$^{r}$, contains 7.4-kb BamH I-Pst I DNA H. *influenzae lsg locus* | [1] |
| pGEM-LOS-5 | Ap$^{r}$, contains 5.5-kb Hind III-Pst I DNA H. *influenzae lsg locus* | [1] |
| pGEM-LOS-10 | Ap$^{r}$, contains ORF *lsgG* | This study |
| pGEM-LOS-12 | Ap$^{r}$, contains ORF *lsgG* | This study |
| pGEM-C-LOS-12 | Ap$^{r}$, contains ORF *lsgG* and ORF *lsgC* | This study |
| pGEM-D-LOS-12 | Ap$^{r}$, contains ORF *lsgG* and ORF *lsgD* | This study |
| pGEM-E-LOS-12 | Ap$^{r}$, contains ORF *lsgG* and ORF *lsgE* | This study |
| pGEM-F-LOS-12 | Ap$^{r}$, contains ORF *lsgG* and ORF *lsgF* | This study |
| pACYC184 | Cm$^{r}$, Tc$^{r}$ | NEB |
| pACYC-F | Cm$^{r}$, contains ORF *lsgF* | This study |
| pACYC-EF | Cm$^{r}$, contains ORF *lsgE* and ORF *lsgF* | This study |

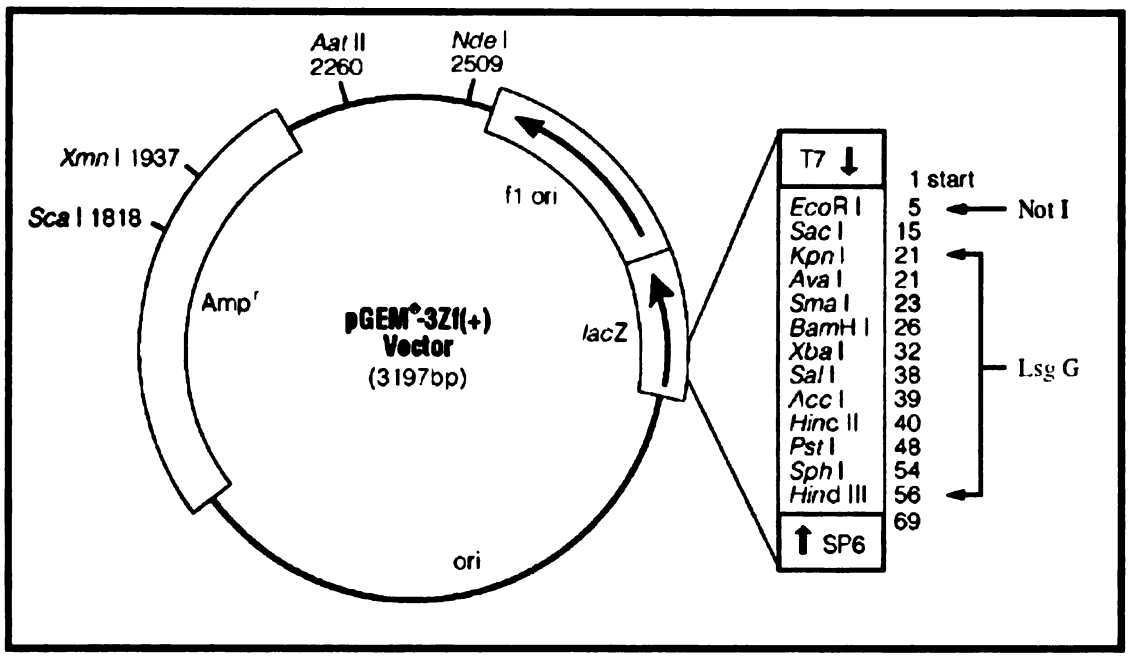
***Bacterial Strains And Vectors:*** The *E. coli* K-12 Strain XL-1 Bluescript was obtained from Stratagene. Strains were routinely cultured at 37°C using LB agar or broth supplemented with the appropriate antibiotics. All vectors used in this study are listed in Table 2-1. The pGEM-LOS-04 and pGEM-LOS-05 vectors used in this study were previously described [1]. Vector pGEM3Z was obtained from Promega, and the vector pACYC184 was obtained from New England Biolabs. Other vectors were constructed as part of this study. All restriction endonucleases were obtained from New England Biolabs (NEB).

***Table 2-2: Primers for Not I Study***

| Primer | Extra BP | Restriction Site | Gene Annealing |
|---|---|---|---|
| EcoR I - Not I Adapter (NEB) | | AATTCgCggCCgCT | |
| *lsgG* Fwd | Cg | ggTACC (Kpn I) | gggAATTCAgggTTTgTT |
| *lsgG* Rev | Tag | AAgCTT (Hind III) | CATgAATCATCTCCTACg |
| *lsgA* Fwd | ACAggAT | gCggCCgC (Not I) | ATCCTTGATATAGAgCgg |
| *lsgA* Rev | ACTAgCT | gCggCCgC (Not I) | CCgCTAATTTACTCgCAT |
| *lsgB* Fwd | CATACAT | gCggCCgC (Not I) | AgTCAATTTgTgggAACg |
| *lsgB* Rev | AgATgAA | gCggCCgC (Not I) | TCCCTgTTTAgCAAgAgC |
| *lsgC* Fwd | TAAgCAg | gCggCCgC (Not I) | CAgCAgAAgATTATCgAg |
| *lsgC* Rev | CAATACT | gCggCCgC (Not I) | gCCTTCTgTTTgTgAAgg |
| *lsgD* Fwd | AgACTAA | gCggCCgC (Not I) | gCgTAAAgCAATTAgTgg |
| *lsgD* Rev | TgATggT | gCggCCgC (Not I) | CgTTgCTCAACgACTACT |
| *lsgE* Fwd | AATTggC | gCggCCgC (Not I) | CATTTTggTTAgCggATg |
| *lsgE* Rev | CAggATT | gCggCCgC (Not I) | gCgTTACTACgCCATCAA |
| *lsgF* Fwd | CgTTAAA | gCggCCgC (Not I) | ggTAgAgCAAgCTCgTAA |
| *lsgF* Rev | CTAgTTg | gCggCCgC (Not I) | gCAACAggTTTggATTCA |

***Combinatorial Ligation System: Construction of pGEM-LOS10:*** To construct the background vector for the combinatorial ligation system, a Not I restriction site and a functional *H. influenzae lsgG* ORF were inserted into the medium copy vector pGEM-3zf+ (Promega). See Figure 2-4A for a schematic. First, primers were designed to amplify the *lsgG* ORF and flanking sequences using the sequence of the *Haemophilus influenzae lsg* locus (NCBI accession number M94855). A Kpn I restriction site and two extra bp to ensure adequate restriction were added to the upstream primer. Likewise, a Hind III site and three extra base pairs were added to the downstream primer (Table 2-2). The primers were synthesized by the UCSF Biomolecular Resource Center (http://www.ucsf.edu/brc/) and PCR of the ORF was performed, using the LOS-04 vector as a template, with the Vent high fidelity thermophilic polymerase (NEB). The PCR product and the vector pGEM3zF+ were each

***Figure 2-4A: Schematic Drawing of PGEM-LOS-10***




Depiction of the vector pGEM-LOS-10. A Not I site allows cloning of random combinations of glycosyltransferase genes. The ORF *lsgG* was cloned into The Hind III and Kpn I sites. The vector also features a Col E1 origin of replication, an ampicillin resistance cassette, and a T7 promoter sequence. Vector graphic courtesy of Promega.

digested with Hind III and Kpn I and purified with Promega Wizard DNA Purification PCR Preps (Wizard PCR preps) to remove DNA oligos and restriction enzymes from the mixture. The concentration of the resultant DNA fragments was determined using the Invitrogen PicoGreen fluorescent DNA quantitation system (PicoGreen). Then the *lsgG* PCR product was ligated into the pGEM3Zf+ vector overnight at 16°C using T4 ligase (NEB). The ligation mixture was transformed into electrocompetent XL-1 Blue (Stratagene) cells using a Micropulser electroporator (Bio-Rad, Hercules, CA). The electroporation cocktail was plated onto LB agar plates supplemented with ampicillin. Colonies were screened by restriction analysis with Kpn I and Hind III and then a sample of plasmid DNA from the successful transformant was digested with EcoR I (NEB).

Next, the EcoR I digested pGEM+*lsgG* vector was incubated with CIP (Roche Applied Science), for 15 min. Concurrently, an aliquot of the EcoR I – Not I adapter (NEB) was incubated with T4 polynucleotide kinase (NEB) for 15 min. The EcoR I digested, CIP treated, pGEM+*lsgG* vector was run through a Wizard PCR Prep to remove enzyme contaminants. Then the phosphorylated EcoR I – Not I adapter was ligated into the EcoR I site of this vector with T4 Ligase overnight at 16°C. The ligation mixture was transformed into electrocompetent XL-1 Blue cells with a Bio-Rad Micropulser electroporator. The electroporation cocktail was plated onto LB agar plates supplemented with ampicillin. Colonies were screened by restriction analysis with Not I and Hind III and a MIDI prep of the successful transformant was performed. The sequence of the successful pGEM-LOS-10 vector was confirmed via sequencing with an Applied Biosystems automated sequencer using fluorescent terminator dye tags at the UCSF Biomolecular Resource Center (http://brc.ucsf.edu).

***Combinatorial Ligation System: PCR of lsgA, B, C, D, E and F and Ligation:*** In order to obtain a combinatorial assortment of 1-3 glycosyltransferase genes in the pGEM-LOS-10 vector, five genes were amplified by PCR, mixed together and ligated into the Not I site of the vector. First, primers were designed to amplify each individual ORF (and flanking sequences) of *lsgA, B, C, D,* and *E* using the sequence of the *H. influenzae lsg* locus. A Not I restriction site was added to the end of each primer as well as seven extra nucleotides to ensure adequate restriction. The sequences of all primers from this study are listed in Table 2-2. The primers were synthesized by the UCSF Biomolecular Resource Center and then PCR of each ORF was performed, using the LOS-04 vector as a template, with the Vent high fidelity thermophilic polymerase. Each of the six PCR products were digested with Not I and each was run though a Wizard PCR prep to remove contaminating oligonucleotides and enzymes. The resulting purified PCR products were quantitated with PicoGreen. A master mixture was created with 10 ng/μl of each gene. Meanwhile, an aliquot of the pGEM-LOS-10 vector was digested with Not I, treated with CIP, and run through a Wizard PCR prep to remove contaminating enzymes. The resulting purified, Not I digested vector was also quantitated using PicoGreen. Several attempts were made to ligate between one and four genes into the Not I site of the pGEM-LOS-10 vector. Aliquots of the master mixture were combined with aliquots of the purified, Not I digested pGEM-LOS-10 vector in the following gene to vector ratios. 100:10, 100:33.3, and 100:50. To each mixture T4 Ligase, buffer and ATP were added. Each was incubated at 16°C overnight, then transformed into electrocompetent XL-1 Blue cells with a Bio-Rad Micropulser electroporator. The

electroporation cocktail was plated onto LB agar plates supplemented with ampicillin. Colonies were screened by restriction analysis with Not I.

***Phosphorylation Limited 3-Gene Ligation:*** In order to enrich the population of 3-gene ligation products, "Phosphorylation Limited 3-Gene Ligation" was attempted. This was done using the same primers designed for the Not I combinatorial ligation system. First, PCR of three *lsg* gene locus ORFs was performed, using the LOS-04 vector as a template, with Vent high fidelity thermophilic polymerase. Each PCR product was digested with Not I. Then, two of the three PCR products were incubated with CIP and dephosphorylated. The remaining PCR product was left with its 5' phosphates intact. Meanwhile, an aliquot of the pGEM-LOS-10 vector was digested with Not I and treated with CIP. All three PCR products and the vector were run through a Wizard PCR prep to remove contaminating enzymes, and quantitated using the PicoGreen system. The 3-gene ligation took place in two steps. First, the three genes were combined such that there was an excess of the two dephosphorylated PCR products, and a limiting concentration of the PCR product with intact phosphates (Figure 2-1). The resulting ligation cocktail was separated by agarose gel electrophoresis, and the band corresponding to three genes was excised and purified using the Qiaex II DNA gel extraction kit (Qiagen). The resulting 3-gene DNA fragment was incubated with T4 Polynucleotide Kinase, run through a Wizard PCR prep and quantitated with PicoGreen. Second, the phosphorylated 3-gene DNA fragment was ligated into the Not I digested, dephosphorylated pGEM-LOS-10 vector overnight at 16°C with T4 ligase. The ligation

mixture was transformed into electrocompetent XL-1 Blue cells with a Bio-Rad Micropulser electroporator. The electroporation cocktail was plated onto LB agar plates supplemented with ampicillin. Colonies were screened by restriction analysis with Not I.

***Hapaxomer System: Construction of pGEM-LOS-12:*** In order to achieve 3-gene ligation using the hapaxoterministic endonuclease Dra III, a background vector with two different Dra III sites and a functional *lsgG* ORF was created. First, primers were designed to amplify the *lsgG* ORF and flanking sequences using the sequence of the *Haemophilus influenzae lsg* locus (NCBI accession number M94855). Pst I, and two different Dra III restriction sites were added to the upstream primer as well as two extra base pairs to ensure adequate restriction. Likewise, a Hind III site and two extra base pairs were added to the downstream primer (Table 2-3). Figure 2-4B is a schematic of pGEM-LOS-12.

The primers were synthesized by Integrated DNA Technologies (http://www.idt.com). PCR of the ORF was performed, using the LOS-04 vector as a template, with the Easy-A high fidelity thermophilic polymerase cocktail (Stratagene). The PCR product and the vector pGEM3z were each digested with Hind III and Pst I. The digested PCR product and vector were then purified and concentrated to 6 $\mu$l using 5 $\mu$g Zymo Clean & Concentrate minicolumn kits (Zymo Labs). Two $\mu$l of each were run on an agarose gel to assess concentration and purity and the *lsgG* PCR product was ligated into the pGEM3z vector using the Rapid DNA Ligation Kit (Roche Applied Research). The ligation was again run through a Zymo kit and then was transformed into

electrocompetent XL-1 Blue (Stratagene) cells using a Bio-Rad MicroPulser electroporator. The electroporation cocktail was plated onto LB agar plates supplemented with ampicillin. The DNA sequence of the successful transformant was screened by restriction analysis, and confirmed via sequencing with an Applied Biosystems automated sequencer using fluorescent terminator dye tags at UCSF's Biomolecular Resource Center (http://brc.ucsf.edu).

***Table 2-3: Primers for Dra III Study***

| Primer | Extra BP | Restriction Site[A] | Gene Annealing |
|---|---|---|---|
| | | CtgCAg (Pst I) CACgCCgTggTAACATCACAgCgTg | |
| *lsgG* Fwd | AA | (Two Dra III sites with spacer (blue)) | gggAATTCAgggTTTgT |
| *lsgG* Rev | Tag | AAgCTT (Hind III) | CATgAATCATCTCCTACg |
| *lsgD* Fwd Pos 1 | CAggATCC | CACgCCgTg (Dra III) | TggCTTACCTATTgTTgC |
| *lsgD* Rev Pos 1 | CAggATCC | CACCTTgTg (Dra III) | TTgCCgATTgTCTgTTCg |
| *lsgE* Fwd Pos 2 | CAggATCC | CACAAggTg (Dra III) | CCTTgAggCAATACTTCg |
| *lsgE* Rev Pos 2 | CAggATCC | CACgTggTg (Dra III) | TTgCCgATTgTCTgTTCg |
| *lsgF* Fwd Pos 3 | CAggATCC | CACCACgTg (Dra III) | ggTAgAgCAAgCTCgTAA |
| *lsgF* Rev Pos 3 | CAggATCC | CACgCTgTg (Dra III) | gCAACAggTTTggATTCA |

A: The designer sticky end for each Dra III site is highlighted in red. See Figure 2-3 for a key.

***Hapaxomer System: Hapaxoterministic 3-Gene Ligation:*** A Hapaxoterministic system was investigated in order to obtain a 3-gene single step ligation. A system of designer 3bp overhangs was adapted [2] that would allow for three cloning segments in a single vector (Figure 2-3). To accomplish this 3-gene ligation, primers were designed to amplify the ORF and flanking sequences for a test set of three genes (Table 2-3). A Dra III site was added to each primer with a designer sticky end corresponding to the desired position of that gene. For example: A Dra III restriction site with the designer sticky end of GCC

was added to the upstream primer and a Dra III site with the designer sticky of TTC was added to the downstream primer for the amplification of the *lsgD* ORF. These ends would place it in position 1 of the 3-gene ligation (Figure 2-3). The primers were synthesized by Integrated DNA Technologies. PCR of each individual ORF was performed, using the LOS-04 vector as a template, with the Easy-A high fidelity thermophilic polymerase cocktail (Stratagene). The three PCR products and the vector pGEM-LOS-12 were each digested with Dra III. The digested PCR product and vector were then purified and concentrated to 6 $\mu$l using 5 $\mu$g Zymo Clean & Concentrate minicolumn kits. Two $\mu$l of each were run on a gel to assess concentration and purity and the three PCR products were ligated into the pGEM-LOS-12 vector using the Rapid DNA Ligation Kit (Roche Applied Research). The ligation was again run through a Zymo kit and transformed into electrocompetent XL-1 Blue (Stratagene) cells using a Bio-Rad MicroPulser electroporator. The electroporation cocktail was plated onto LB agar plates supplemented with ampicillin. The DNA sequence of the successful transformant was screened by restriction analysis.

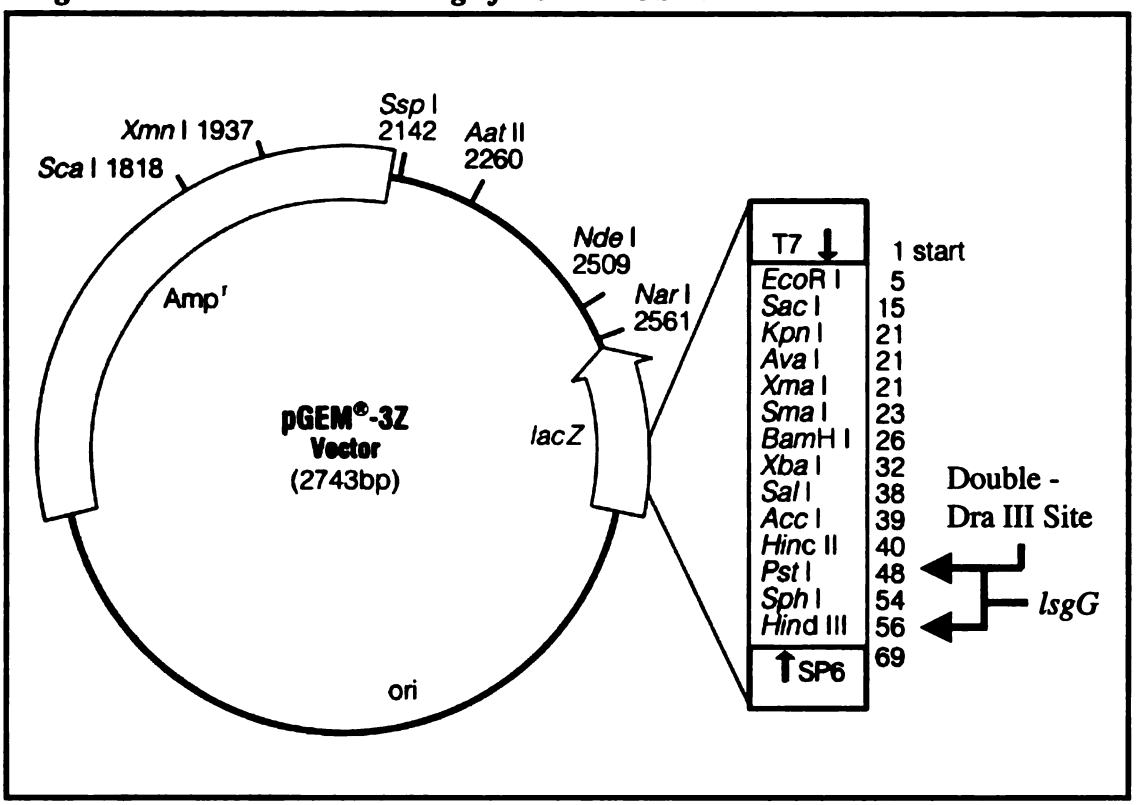
***Multiplasmid System: General.*** In another strategy for the expression of 1-3 genes in *E. coli* K-12, a multi-vector system was designed. pGEM-LOS-12 was one suitable background vector. The vector pACYC184 was selected as the second plasmid because it has an origin of replication that is compatible with the Col E1 origin of replication on pGEM-LOS-12 and has a different antibiotic resistance gene. The general plan for this system was to create four single gene + pGEM-LOS-12 vectors and a smaller set of

pACYC184 vectors. Combining each pGEM vector with each pACYC184 vector would allow the genes to be mixed and matched without extensive cloning steps.

To decrease background in the ligation reactions, a suicide cloning method was implemented. In suicide cloning, two different restriction sites with compatible sticky ends are employed [14,15]. Upon successful ligation there is no longer a restriction site for either endonuclease. Digesting the mixture after ligation with the endonuclease originally used to linearize the vector, reduces background re-circularized vector. In this study, the compatible endonucleases used were Xho I and Sal I.

***Multiplasmid System: Construction of pGEM-LOS-12 C, D, E, and F.*** In order to create the single gene + *lsgG* vectors, primers were designed to amplify the relevant ORF (*lsgC*, *lsgD*, *lsgE* and *lsgF*, respectively), and flanking sequences, using the sequence of the *Haemophilus influenzae lsg* locus. Xho I restriction sites were added to the upstream and downstream primers in addition to three extra base pairs to ensure adequate restriction (Table 2-4). Primers were synthesized by Integrated DNA Technologies. PCR of the ORF was performed, using the LOS-04 vector as a template, with the Easy-A high fidelity thermophilic polymerase cocktail (Stratagene). PCR products were digested with Xho I while the pGEM-LOS-12 vector was digested with Sal I (NEB) and treated with CIP. The digested PCR product and vector were purified and concentrated to 6 $\mu$l using

***Figure 2-4B: Schematic Drawing of PGEM-LOS-12***




Depiction of the vector pGEM-LOS-12. Two Dra III sites allow two different "designer sticky ends" for each end of the cloned inserts. The ORF *lsgG* was cloned into the Hind III and Pst I sites. The vector also features a Col E1 origin of replication, an ampicillin resistance cassette, and a T7 promoter sequence. Vector graphic courtesy of Promega.

Zymo kits. Two $\mu$l of each were run on a gel to assess concentration and purity, and the Xho I digested PCR product was ligated into the Sal I site of the pGEM-LOS-12 vector using the Rapid DNA Ligation Kit (Roche Applied Research). The ligation was digested for one hour with Sal I to reduce background recircularized vector. It was then run through a Zymo kit and transformed into electrocompetent XL-1 Blue cells using a Bio-Rad MicroPulser electroporator. The electroporation cocktail was plated onto LB agar plates supplemented with ampicillin. The DNA sequence of the successful transformant was screened by restriction analysis, and confirmed via sequencing with an Applied Biosystems automated sequencer using fluorescent terminator dye tags at the UCSF Biomolecular Resource Center.

***Table 2-4: Primers for the Multiplasmid System***

| Primer | Extra BP | Restriction Site | Gene Annealing |
|---|---|---|---|
| *lsgC* Fwd | ACA | AgaTCT (Xho I) | CAgCAgAAgATTATCgAg |
| *lsgC* Rev | ACA | AgaTCT (Xho I) | gCCTTCTgTTTgTgAAgg |
| *lsgD* Fwd | TCA | AgaTCT (Xho I) | gCgTAAAgCAATTAgTgg |
| *lsgD* Rev | TCA | AgaTCT (Xho I) | CgTTgCTCAACgACTACT |
| *lsgE* Fwd | ACA | AgaTCT (Xho I) | CATTTTggTTAgCggATg |
| *lsgE* Rev | ACA | AgaTCT (Xho I) | gCgTTACTACgCCATCAA |
| *lsgF* Fwd | TCA | AgaTCT (Xho I) | ggTAgAgCAAgCTCgTAA |
| | TCA | AgaTCT (Xho I) | gCAACAggTTTggATTCA |

***Multiplasmid System: Construction pACYC184-lsgF.*** The first pACYC vector required was pACYC184-*lsgF*. Utilizing the primers designed for the pGEM-F-LOS-12 vector, a fresh PCR amplification of the *lsgF* ORF was performed. The PCR products were

cloned into the Sal I site of the pACYC184 vector using the suicide cloning techniques described in the previous section, with the following exception. The electroporation cocktail was plated onto LB agar plates supplemented with chloramphenicol. The DNA sequence of the successful transformant was screened by restriction analysis, and confirmed via sequencing with an Applied Biosystems automated sequencer using fluorescent terminator dye tags at the UCSF Biomolecular Resource Center.

***Multiplasmid System: Construction pACYC184-lsgEF.*** The only other pACYC vector required to complete the study of the four glycosyltranferases *lsgCDEF* from the *H. influenzae lsg* gene locus was pACYC184-*lsgEF*. Using the upstream primer designed for the pGEM-E-LOS-12 vector and the downstream primer designed for the pGEM-F-LOS-12 vector, a fresh PCR amplification of the *lsgEF* ORFs was performed. This PCR product was cloned into the Sal I site of pACYC184 vector using the suicide cloning techniques previously described. The electroporation cocktail was plated onto LB agar plates supplemented with chloramphenicol. The DNA sequence of the successful transformant was screened by restriction analysis, and confirmed via sequencing with an Applied Biosystems automated sequencer using fluorescent terminator dye tags at the UCSF Biomolecular Resource Center.

***Multiplasmid System: Creating Two Vector Transformants:*** In order to create two vector strains with combinations of a pACYC184 vector and a pGEM vector, *E. coli*

containing the relevant pACYC184 vector (pACYC-F or pACYC-EF, respectively) were cultured overnight in 20 ml of chloramphenicol- supplemented LB. The following day, 5 ml of those cultures were added to 60 ml LB:chloramphenicol and the bacteria were grown with shaking at 37°C until the bacteria reached an OD600 of 0.6. The bacteria were harvested by centrifugation at 5400 x *g* in a Jouan MR1812 centrifuge. The pellet was washed three times with 50 ml of ice-cold sterile deionized water and then resolubilized in 160 $\mu$l of ice-cold sterile deionized water. Forty $\mu$l aliquots of the electrocompetent pACYC vector bacteria were transformed with their respective LOS-12 vectors using a Bio-Rad MicroPulser electroporator. The electroporation cocktail was plated onto LB agar plates supplemented with ampicillin and chloramphenicol. Qiagen mini-plasmid preps of the successful transformants were then screened via PCR for the presence of their respective vectors. Table 2-5 lists all of the single-vector and double vector transformants created in this study.

***Table 2-5: Multiplasmid System E. coli Transformants***

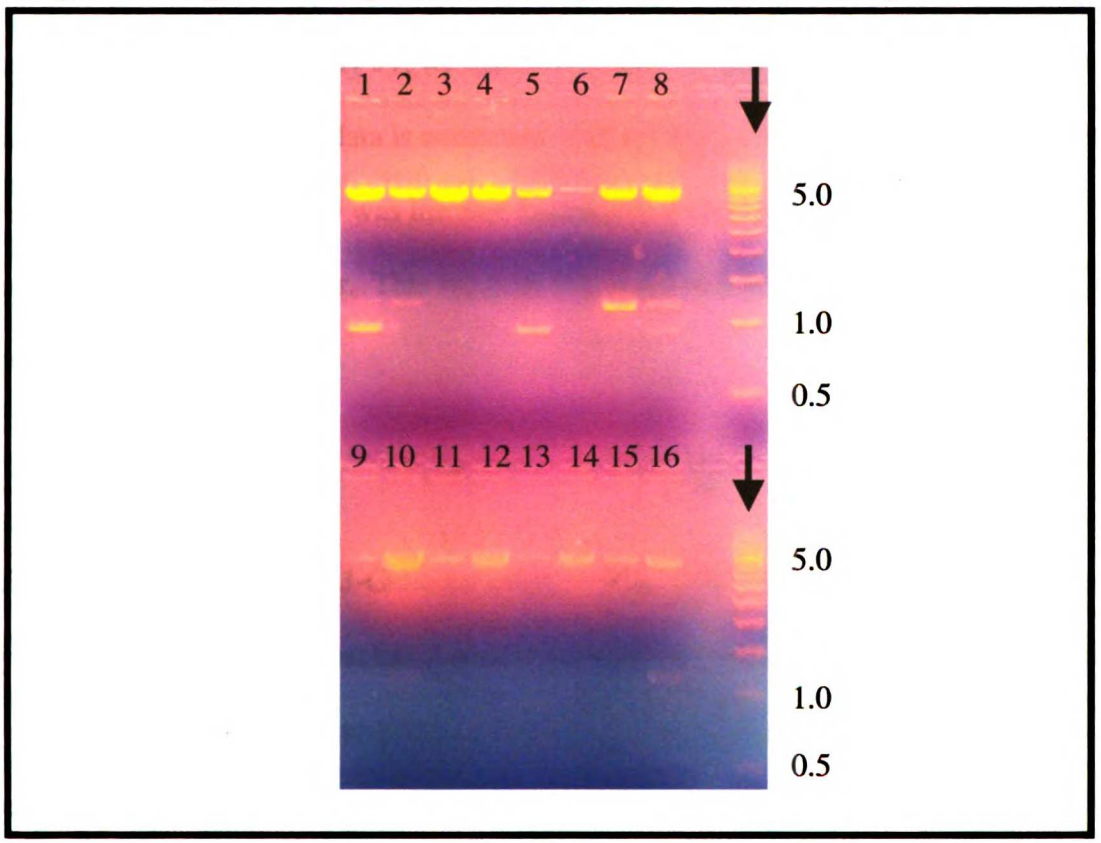
| Transformant | Vector 01 | Vector 02 | Relevant Hib ORFs |
|---|---|---|---|
| LOS-12 | pGEM-LOS-12 | | *lsgG* |
| CG | pGEM-C-LOS-12 | | *lsgC, lsgG* |
| DG | pGEM-D-LOS-12 | | *lsgD, lsgG* |
| EG | pGEM-E-LOS-12 | | *lsgE, lsgG* |
| FG | pGEM-F-LOS-12 | | *lsgF, lsgG* |
| CFG | pGEM-C-LOS-12 | pACYC-F | *lsgC, lsgF, lsgG* |
| DFG | pGEM-D-LOS-12 | pACYC-F | *lsgD, lsgF, lsgG* |
| EFG | pGEM-E-LOS-12 | pACYC-F | *lsgE, lsgF, lsgG* |
| CEFG | pGEM-C-LOS-12 | pACYC-EF | *lsgC, lsgE, lsgF, lsgG* |
| DEFG | pGEM-D-LOS-12 | pACYC-EF | *lsgD, lsgE, lsgF, lsgG* |

## RESULTS

***Overview:*** In general we found that the techniques used to attempt multiple-gene ligations were not sufficiently effective at the time of this study. The only method robust and versatile enough for our purposes was the multiplasmid method. Single – gene ligations are well within the capabilities of T4 ligase, and once these vectors were successfully transfected, they were quite stable as well.

***Combinatorial Ligation:*** Once the pGEM-LOS-10 vector was successfully created, several attempts were made to ligate a mixture of genes into the vector at various insert:vector ratios. It was first thought that a genes:vector ratio of 3:1 would result in approximately 3 inserts per vector. However, several attempts to achieve this number of genes/vector routinely resulted in unsuccessful transformations in which the colonies contained re-circularized vector or single-gene vectors upon restriction analysis with Not I. The ratio of genes:vector was increased to 10:1 for the next round of combinatorial ligations. In screening the colonies that resulted from the transformation of this ligation cocktail into *E. coli*, we began to observe small numbers of multi-gene vectors. Figure 2-5 is an image of the DNA gel of Not I digested DNA from sixteen of these colonies. Eight of the sixteen colonies only contain the ~4.5 kb band corresponding to Not I digested pGEM LOS-10. Of the eight colonies that do have inserts, only colony #8 is immediately apparent as having more than one gene. Colonies 1, 2, 5, 7, 9, 10 and 16 may have more than one gene, as some of the genes are the same size (roughly half are

***Figure 2-5: Agarose Gel of Plasmid DNA from Combinatorial Ligation***

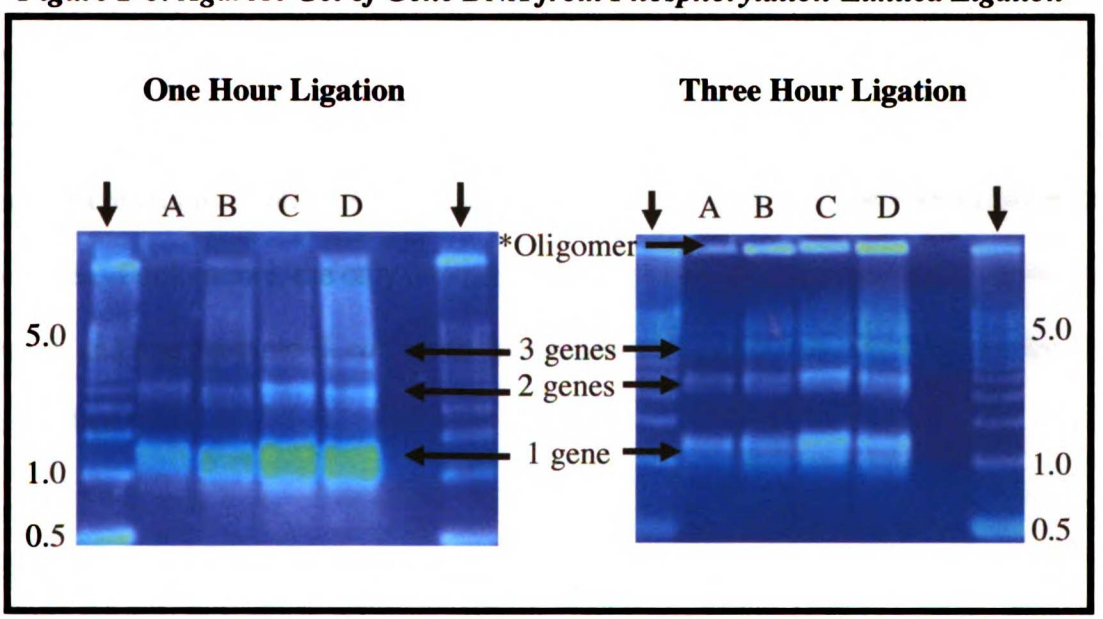



1-10.) Plasmid DNA from successful transformants of the 10:1 genes:vector ratio combinatorial ligation. The plasmid DNA was extracted from individual colonies and digested with Not I. ↓ = Bio-Rad 1kb Lane Marker

slightly larger than 1 kb, half are slightly smaller than 1 kb). Of the eight colonies surveyed, colony #7 was the only colony to display an altered phenotype. These results were promising, as we had demonstrated that at least one of these genes is capable of functioning outside the *lsg* operon. However, out of sixteen colonies, one with an interesting phenotype is not indicative of the level of robustness required to test all possible combinations of these genes. By sequencing the plasmid DNA from combinatorial ligation colony #7, we determined that the gene responsible for the altered phenotype was *lsgF*. This data is consistent with previous experiments [1]. In addition, we found that the *lsgF* ORF was inserted backwards with respect to the T7 promoter present on the pGEM vector. This suggests that there may be promoter sequences upstream of the ORF itself.

***Phosphorylation Limited 3-Gene Ligation:*** It was our intention to ultimately use this system to bring about all possible 3-gene combinations with the five glycosyltransferase genes. However, as a pilot project I only attempted to generate the 3-gene ligation between *lsgD, E,* and *F*. Several attempts were made to achieve step one of the two-step cloning process. A ligation cocktail consisting of 10ng of phosphorylated *lsgE*, 100ng of dephosphorylated *lsgD* and 100ng of dephosphorylated *lsgF* Not I digested PCR products was ligated overnight at 16°C. Agarose Gel electrophoresis of this ligation cocktail however, revealed a problem. The ligation reactions had resulted in a DNA band so large it hardly traveled out of the well. However there were still DNA bands present in the ~1.0 kb region of the gel. Between the two extremes, the gel was relatively empty. Several

***Figure 2-6: Agarose Gel of Gene DNA from Phosphorylation-Limited Ligation***




Agarose gel electrophoresis of 3-gene phosphorylation-limited ligation. A.)10 ng *lsgD*, 1 ng *lsgE*, 10 ng *lsgF*, B.)25 ng *lsgD*, 2.5 ng *lsgE*, 25 ng *lsgF*. C.)50 ng *lsgD*, 5 ng *lsgE*, 50 ng *lsgF*, d.)100 ng *lsgD*, 10 ng *lsgE*, 100 ng *lsgF*. ↓ = Bio-Rad 1kb Lane Marker *Note the presence of large oligomer in just 3 hr of ligase reaction.

attempts were made to control the ligation reaction, however DNA ligase is difficult to regulate. Figure 2-6 is a SYBR Gold (Invitrogen) stained agarose gel of gene-ligation product DNA resulting from a range of concentrations and ligation times. This gel represents the best out of over 20 attempts at varying ligation time, DNA concentrations and ratios, and ligation temperatures. SYBR gold is capable of visualizing much lower concentrations of DNA than ethidium bromide. The appearance of the gel indicates that a one-hour ligation of the three genes barely results in detectable levels of 3-gene ligation products. It is also apparent that the giant oligomer had already formed in just two hours, yet the amount of three-gene ligation products was barely visible. It appeared that the ease of combining ligation products with the system will not be as optimal as this project requires. Furthermore, the only 3-gene product we were able to successfully ligate into a vector and transfect into *E. coli* using this system was found to contain three copies of the *lsgE* ORF.

***Hapaxomer System: Hapaxoterministic 3-Gene Ligation:*** Several attempts were made to accomplish this 3-gene ligation under exceedingly stringent reaction conditions and ultimately every colony screened was a recircularized vector or some amalgamation of vector and a small insert. Extensive attempts were made to remove small bits of oligonucleotides that could have re-ligated the vector. This technique was performed at and the following gene:vector ratios: 1:1:1:10; 1:1:1:5; 1:1:1:1; 5:5:5:1; and 10:10:10:1, at low ng, mid ng and low μg concentrations. No successful multi-gene transformant was ever isolated using this system.

***Multiplasmid System:*** The multiplasmid system was successfully used for the purposes of expressing combinations of up to three different genes in *E. coli*. Each single and double gene vector in Table 2-5 was synthesized with little difficulty. DNA sequencing confirmed each gene or genes in their respective vectors. It was then very simple to express the vectors in different combinations. As long as the vectors had compatible origins of replication, and different antibiotic resistance genes, their products were expressed successfully. This technique enabled us to decode the phenotypes of the *H. influenzae lsg* gene locus glycosyltransferases *lsg C, D, E, and F* [16]. This system is robust and versatile and can be adapted to screening larger groups of lipopolysaccharide synthesis glycosyltransferase genes in the future.

## DISCUSSION

Four methods for combining large segments of DNA have been reviewed for their ease of use, robustness, and versatility in a multi-gene expression system. The goal of this study was to engineer a method of expressing one to three 1 kb genes per *E. coli* colony. Both combinatorial methods and directed methods were explored. Criteria for a successful combinatorial method would be that there were a high number of two to three gene containing colonies that represented a range of different gene combinations. Criteria for a successful directed methodology would be the ability to engineer a system in which one to three gene containing colonies could be mixed and matched in a facile

manner. Three of the four methods profiled, however, relied on DNA ligase, the enzyme typically used to covalently attach two segments of DNA. These three methods, in which DNA ligase was required to ligate more than one gene into one vector at a time, did not meet the stated criteria of a robust and versatile expression system.

In the case of the combinatorial ligation method, it was difficult to obtain vectors with more than one insert that would stably transfect *E. coli.* Only a single phenotypic variant was discovered among a test sampling of 16 colonies, and 16 colonies were not enough to cover the range of combinations we required. Combinatorial Not I ligation did not meet the criteria for a successful expression system.

The phosphorylation limited 3-gene ligation strategy did not meet the criteria either. However, separating the experiment out into two steps did expose a key problem with DNA ligase. Because we were able to visualize the first reaction step with a DNA gel, it was apparent that the DNA ligation reaction would be difficult to control by varying reaction parameters such as time, and concentration of DNA. A two hour ligation showed few discrete steps between one/two gene bands and a giant oligomer, not a ladder pattern of gene multiples as might be expected. In one hour, the same reaction had not even progressed to the 3-gene stage. Several other reaction times and concentrations were attempted with similarly unremarkable results. DNA ligase was too difficult to regulate for the purposes of this study.

The "Hapaxomer" system should have transferred the responsibility for generating 3-gene constructs from DNA ligase to the annealing step and the sequence of the 3 bp overhangs (Figure 2-3). However, no successful transformants ever resulted from this system. Fifteen different sets of reaction conditions were attempted with no

successful result. There is something fundamentally wrong when this system is used with 1kb inserts. It is possible that this technique worked for Leibler and coworkers [2], because their inserts were smaller. The ratio between the size of the insert and the length of the sticky-ends of the overlaps may be essential. It is possible that the energy gained from the three base pair overlaps annealing with each other between each gene may not be enough to compensate for the large entropy loss inherent in organizing three free linear 1 kb inserts and a free, linear 4.5 kb vector into a single, circular 7.5 kb plasmid, with no free ends. Recent studies in total gene synthesis suggest that overlaps of 10-20 bp can overcome this problem [17-19]. Adapting this system to 10 bp overlaps would require that we re-engineer the system completely, as no endonucleases have yet been discovered that allow custom design of 10 bp overlaps. At the time of this study, that would have required investigation of total-gene synthesis methods, and would not have been cost-effective for our purposes. Ultimately, the "Hapaxomer" method did not meet the criteria for a successful expression system.

The "Multiplasmid" method did meet our criteria for a successful expression system. Synthesis of one and two-gene vectors was reasonably facile and was accomplished in parallel. Each gene of interest was cloned alongside the necessary *lsgG* gene and an ampicillin resistance gene in four different pGEM vectors. The pGEM vectors were co-transfected alongside two different pACYC184 vectors respectively. Altogether, E. *coli* was transfected with 10 different *lsg* gene combinations, and they were evaluated for phenotypic variation [16]. This system suited our immediate purposes, and could be expanded upon for the evaluation of larger sets of glycosyltransferase genes as well.

Ultimately, this system could be put to use as a screen for novel glycosyltransferase gene function. A library of glycosyltransferase gene-containing background pGEM-LOS-12 vectors could be assembled, each capable of synthesizing a known lipopolysaccharide phenotype. Glycosyltransferase genes of unknown function could be subcloned into pACYC184 vectors and transfected into *E. coli*. Each pACYC184 clone could be made electrocompetent and transformed with each vector from the background pGEM-LOS-12 vector library. Each of the resulting colonies could then be screened for LPS phenotypic variation. This system could easily be worked into a high-throughput system for determining the functions of novel microbial lipopolysaccharide glycosyltransferases.

## REFERENCES


1. Phillips NJ, Miller TJ, Engstrom JJ, Melaugh W, McLaughlin R, Apicella MA, Gibson BW: **Characterization of chimeric lipopolysaccharides from Escherichia coli strain JM109 transformed with lipooligosaccharide synthesis genes (lsg) from Haemophilus influenzae**. *J Biol Chem* 2000, **275**:4747-4758.
2. Guet CC, Elowitz MB, Hsing W, Leibler S: **Combinatorial synthesis of genetic networks**. *Science* 2002, **296**:1466-1470.
3. Berger SL: **Expanding the potential of restriction endonucleases: use of hapaxoterministic enzymes**. *Anal Biochem* 1994, **222**:1-8.
4. Preston A, Mandrell RE, Gibson BW, Apicella MA: **The lipooligosaccharides of pathogenic gram-negative bacteria**. *Crit Rev Microbiol* 1996, **22**:139-180.
5. Varki A: **Biological roles of oligosaccharides: all of the theories are correct**. *Glycobiology* 1993, **3**:97-130.
6. Zwahlen A, Rubin LG, Moxon ER: **Contribution of lipopolysaccharide to pathogenicity of Haemophilus influenzae: comparative virulence of genetically-related strains in rats**. *Microb Pathog* 1986, **1**:465-473.
7. Kimura A, Hansen EJ: **Antigenic and phenotypic variations of Haemophilus influenzae type b lipopolysaccharide and their relationship to virulence**. *Infect Immun* 1986, **51**:69-79.
8. Coutinho PM, Deleury E, Davies GJ, Henrissat B: **An evolving hierarchical family classification for glycosyltransferases**. *J Mol Biol* 2003, **328**:307-317.

9. Breton C, Mucha J, Jeanneau C: **Structural and functional features of glycosyltransferases**. *Biochimie* 2001, **83**:713-718.
10. Phillips NJ, John CM, Reinders LG, Gibson BW, Apicella MA, Griffiss JM: **Structural models for the cell surface lipooligosaccharides of Neisseria gonorrhoeae and Haemophilus influenzae**. *Biomed Environ Mass Spectrom* 1990, **19**:731-745.
11. Carreras CW, Santi DV: **Engineering of modular polyketide synthases to produce novel polyketides**. *Curr Opin Biotechnol* 1998, **9**:403-411.
12. Stemmer WPC: **Rapid evolution of a protein in vitro by DNA shuffling**. *Nature* 1994, **370**:389-391.
13. Hertweck C: **The multiplasmid approach: a new perspective for combinatorial biosynthesis**. *Chembiochem* 2000, **1**:103-106.
14. Stull B: **Personal Correspondence and His Own Molecular Biology Protocols**. Edited by: UCSF; 2004.
15. Sambrook J, Russell DW: *Molecular cloning : a laboratory manual* edn 3rd. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press; 2001.
16. Johansen E, Gibson BW: **Functional Exploration of the *lsg* gene locus glycosyltransferases *lsgC, D, E, and F*, through chimeric expression in Eschericia coli K-12.** *In process* 2006.
17. Menzella HG, Reid R, Carney JR, Chandran SS, Reisinger SJ, Patel KG, Hopwood DA, Santi DV: **Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes**. *Nat Biotechnol* 2005, **23**:1171-1176.
18. Kodumal SJ, Patel KG, Reid R, Menzella HG, Welch M, Santi DV: **Total synthesis of long DNA sequences: synthesis of a contiguous 32-kb polyketide synthase gene cluster**. *Proc Natl Acad Sci U S A* 2004, **101**:15573-15578.
19. Smith HO, Hutchison CA, 3rd, Pfannkoch C, Venter JC: **Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides**. *Proc Natl Acad Sci U S A* 2003, **100**:15440-15445.

# CHAPTER 3.

# FUNCTIONAL EXPLORATION OF THE *Haemophilus influenzae lsg* LOCUS GLYCOSYLTRANSFERASES *lsgC, lsgD, lsgE, and lsgF*, VIA CHIMERIC EXPRESSION IN *Escherichia coli* K-12

## ABSTRACT

In the current study, I adapted a chimeric expression system to explore the synthetic properties (specificity and selectivity) of four glycosyltransferase genes from the *Haemophilus influenzae lsg* gene locus: *lsgC, D, E,* and *F* in *Escherichia coli* (*E. coli)* K-12. This expression system is unique in that LPS O-antigen production that was previously knocked out in *E. coli* has been restored by the addition of the *H. influenzae lsgG* gene. HF-treated, O-deacylated LPS and HF-treated free oligosaccharide (OS) from 10 *E. coli* strains were prepared from strains that contained either an individual gene or a strategic combination of *lsg* glycosyltransferase genes from the study set paired with *lsgG.* The LOS samples were analyzed by MALDI-TOF, ELISA and linkage analysis in order to determine the gene product(s) responsible for each monosaccharide extension, and, to a limited degree, the degree of selectivity these glycosyltransferases might exhibit with respect to the nascent oligosaccharide chain.

The LsgG gene product rescues O-antigen synthesis resulting in the addition an N-acetyl-glucosamine (GlcNAc) to the 7-position of the terminal non-reducing heptose of the *E. coli* core lipopolysaccharide (LPS) [1]. Analysis of four chimeric *E. coli* transformants containing *lsgCG, lsgDG, lsgEG,* or *lsgFG* revealed that the LsgF gene product was the only glycosyltransferase capable of adding a galactose (Gal) to the

terminal GlcNAc of this GlcNAc-Core LPS structure. Analysis of transformants containing *lsgCFG, lsgDFG, and lsgEFG* revealed that LsgE was the only gene product capable of adding GlcNAc to the terminal Gal of the Gal-GlcNAc-Core LPS structure synthesized by LsgFG. Once this terminal GlcNAc was in place, the LsgF gene product (also present in this transformant) was able to add an additional Gal to this structure as well. Analysis of transformants containing *lsgCEFG* and *lsgDEFG* gene combinations demonstrated that the LsgD gene product competes effectively with LsgF for the extension of the GlcNAc-Gal-GlcNAc-core LPS structure by one Gal.

## INTRODUCTION

Glycoproteins, glycolipids, and free oligosaccharides play pivotal roles in a wide range of biological processes, including cell-cell recognition, inflammation, immunological response, cancer metastasis, and bacterial and viral infection [2-4]. Despite the multitude of carbohydrate structures in nature, the role of saccharide function has not been investigated as extensively as nucleic acids and proteins. Biologists simply do not have the access to suitable amounts of synthetic complex carbohydrates with which to initiate in-depth studies. Automated solid-phase methodologies have revolutionized protein and nucleic acid science, but decades of synthetic research have yet to provide robust methods for automated oligosaccharide synthesis [4,5].

Enzymatic synthesis, chemoenzymatic synthesis, and bacterial metabolic pathway engineering have recently emerged as powerful methods for the large-scale synthesis of complex carbohydrates [5,6]. These methods make use of glycosyltransferases, the enzymes responsible for catalyzing the formation of glycosidic bonds in nature, to accomplish large-scale synthesis of carbohydrates. The limited body of knowledge regarding the synthetic properties of these enzymes is a central obstacle to the success of these techniques.

Previously, Phillips and colleagues [1] reported the characterization of chimeric lipopolysaccharides (LPS) from *E. coli* K-12 transformed with lipooligosaccharide synthesis genes (*lsg*) from capsular *Haemophilus influenzae* type b (Hib). Capsular strains of Hib are responsible for invasive infections in humans and use surface lipooligosaccharides (LOS), the major component of their outer membrane, to express

glycoforms that mimic host structures, and may allow the organism to evade or manipulate host defenses [7-11]. Efforts to correlate LOS structures with specific biological functions are hindered by a high degree of heterogeneity and variability of Hib LOS [9,11-13]. Expression of Hib glycoforms in *E. coli* K-12 provides a unique opportunity to facilitate the elucidation of LOS biosynthesis in a simplified system [1,14,15].

The chimeric LPS study characterized extended LPS from three *E. coli* K-12 strains. Each strain was transformed with vectors designated LOS-04, LOS-05 and LOS-07, respectively [1]. It was determined that the chimeric LPS resulting from these transformations consist of the complete *E. coli* LPS core structure, glycosylated on the 7-position of the non-reducing terminal branch heptose of the *E. coli* core LPS with oligosaccharides from Hib. Furthermore, the expression of exogenous LPS was dependent on activation of the *E. coli wecA* gene by the LsgG gene product (present in each successful transformant). The study also characterized these LPS by matrix-assisted laser desorption ionization, electrospray ionization, and tandem mass spectrometry analyses, as well as composition and linkage analyses. A brief summary of structural conclusions from these analyses is featured in Table 3-1.

***Table 3-1: Chimeric oligosaccharides from Phillips et al [1].***

| Phillips et al. Transformant | Relevant Hib ORF | Chimeric Oligosaccharide |
|---|---|---|
| PGEM-LOS-07 | *lsgF, G* | Gal1–3 GlcNAc→CORE |
| PGEM-LOS-05 | *lsgC, D, E, F, G* | Gal1–4 GlcNAc1–3 Gal1–3 GlcNAc→CORE |
| PGEM-LOS-04 | *lsgA, B, C, D, E, F, G* | GlcNAc/Gal1–6 Gal1–4 GlcNAc1–3 Gal1–3 GlcNAc→CORE |

In the current study, I adapted this chimeric expression system to explore the synthetic properties (specificity and selectivity) of the four glycosyltransferase genes from the LOS-05 and LOS-07 constructs: *lsgC, D, E,* and *F*. Each gene was expressed individually with *lsgG,* as well as in combinations of genes that had the potential for sequential carbohydrate addition. In this sequential manner, each step in which the LOS is extended will identify the gene responsible for the sugar addition(s). Comparing these strains with transformants displaying no LPS extension will yield a small array of specificity data for this subset of four Hib glycosyltransferases, which have been studied extensively, yet still elude functional assignment [1,9,11,14,16-18].

The other two genes from the Hib *lsg* gene locus, *lsgA* and *B*, have been implicated as an LPS biosynthesis transporter and a sialyltransferase, respectively and are the focus of another study (Johansen and Zaleski; see Chapter 5.).

In an effort to characterize the synthetic properties of the LsgC,D,E, and F gene products for carbohydrate synthesis and better understand Hib LOS biosynthesis, I report here the partial characterization of the LPS from *E. coli* strain XL-1 transformed with ten different combinations of genes from the *H. influenzae lsg* locus using composition and linkage analysis, ELISA analysis and mass spectrometry.

## MATERIALS AND METHODS

***Bacterial Strains And Vectors:*** The *E. coli* K-12 Strain XL-1 Bluescript was obtained from Stratagene. Strains were routinely cultured at 37°C using LB agar or broth

supplemented with the appropriate antibiotics. All vectors used in this study are listed in Table 3-2. The pGEM-LOS-04 and pGEM-LOS-05 vectors used in this study were previously described [1,8,14,15]. The vector pGEM3Z was obtained from Promega, and the vector pACYC184 was obtained from New England Biolabs. Other vectors were constructed as part of this study. All enzymes were from New England Biolabs (NEB) unless otherwise specified. Standard cloning procedures [19] and suicide cloning methods [20] were used for the construction of the vectors.

***Table 3-2: Bacterial Strains and Vectors***

| Strain/Plasmid | Relevant Characteristics | Ref./Source |
|---|---|---|
| *E. coli* | | |
| XL-1 Bluescript | *recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F'*proAB lacI*$^q$*ZDM15 Tn10* (Tet$^r$)] | Stratagene |
| Plasmids | | |
| pGEM3Z | Ap$^r$ | Promega Biotech |
| pGEM-LOS-4 | Ap$^r$, contains 7.4-kb BamHI-Pst I DNA H. *influenzae lsg locus* | [1] |
| pGEM-LOS-5 | Ap$^r$, contains 5.5-kb Hind III-Pst I DNA H. *influenzae lsg locus* | [1] |
| pGEM-LOS-12 | Ap$^r$, contains ORF *lsgG* | This study |
| pGEM-C-LOS-12 | Ap$^r$, contains ORF *lsgG* and ORF *lsgC* | This study |
| pGEM-D-LOS-12 | Ap$^r$, contains ORF *lsgG* and ORF *lsgD* | This study |
| pGEM-E-LOS-12 | Ap$^r$, contains ORF *lsgG* and ORF *lsgE* | This study |
| pGEM-F-LOS-12 | Ap$^r$, contains ORF *lsgG* and ORF *lsgF* | This study |
| pACYC184 | Cm$^r$, Tc$^r$ | New England Biolabs |
| pACYC-F | Cm$^r$, contains ORF *lsgF* | This study |
| pACYC-EF | Cm$^r$, contains ORF *lsgE* and ORF *lsgF* | This study |

***Construction of pGEM-LOS-12:*** Primers were designed to amplify the *lsgG* ORF and flanking sequences using the sequence of the *Haemophilus influenzae lsg* locus (NCBI accession number M94855). Pst I and two Dra III restriction sites were added to the upstream primer as well as two extra base pairs to ensure adequate restriction. Likewise,

a Hind III site and two extra base pairs were added to the downstream primer (Table 3-3). The primers were synthesized by Integrated DNA Technologies (http://www.idt.com) and PCR of the ORF was performed, using the LOS-04 vector as a template, with the Easy-A high fidelity thermophilic polymerase cocktail (Stratagene). The PCR product and the vector pGEM3z were each digested with Hind III and Pst I. The digested PCR product and vector were then purified and concentrated to 6 $\mu$l using 5 $\mu$g Zymo Clean & Concentrate minicolumn kits (Zymo Labs: http://www.zymoresearch.com/). Two $\mu$l of the purified vector and PCR product were run on a gel to assess concentration and purity and then the *lsgG* PCR product was ligated into the pGEM3z vector using the Rapid DNA Ligation Kit (Roche Applied Research). The ligation was again run through a Zymo kit and then was transformed into electrocompetent XL-1 Blue (Stratagene) cells using a Bio-Rad MicroPulser electroporator. The electroporation cocktail was plated onto LB agar plates supplemented with ampicillin. The DNA sequence of the successful transformant was screened by restriction analysis, and confirmed via sequencing with an Applied Biosystems automated sequencer using fluorescent terminator dye tags at the UCSF Biomolecular Resource Center (http://brc.ucsf.edu).

***Table 3-3: Primers used in cloning of the lsg genes***

| ORF | Upstream Primer | Downstream Primer |
|---|---|---|
| *lsgC* | CAgCAgAAgATTATCgAg | gCCTTCTgTTTgTgAAgg |
| *lsgD* | GCgTAAAgCAATTAgTgg | CgTTgCTCAACgACTACT |
| *lsgE* | CATTTTggTTAgCggATg | gCgTTACTACgCCATCAA |
| *lsgF* | GgTAgAgCAAgCTCgTAA | gCAACAggTTTggATTCA |
| *lsgG* | GggAATTCAgggTTTgTT | CATgAATCATCTCCTACg |

***Construction of pGEM-C-LOS-12, pGEM-D-LOS-12, pGEM-E-LOS-12, and pGEM-F-LOS-12:*** Primers were designed to amplify the relevant ORF (*lsgC*, *lsgD*, *lsgE* and *lsgF*, respectively), and flanking sequences, using the sequence of the *Haemophilus influenzae lsg* locus (NCBI accession number M94855). Xho I restriction sites were added to the upstream and downstream primers as well as three extra base pairs to ensure adequate restriction (Table 3-3). The primers were synthesized by Integrated DNA Technologies and PCR of the ORF was performed, using the LOS-04 vector as a template, with the Easy-A high fidelity thermophilic polymerase cocktail (Stratagene). The PCR product was digested with Xho I while the pGEM-LOS-12 vector was digested with Sal I. The digested PCR product and vector were then purified and concentrated to 6 $\mu$l using Zymo kits. Two $\mu$l of each were run on a gel to assess concentration and purity and then the Xho I digested PCR product was ligated into the Sal I site of the pGEM-LOS-12 vector using the Rapid DNA Ligation Kit (Roche Applied Research). The ligation was digested for one *hour* with Sal I to reduce background recircularized vector. It was then run through a Zymo kit and transformed into electrocompetent XL-1 Blue cells using a Bio-Rad MicroPulser electroporator. The electroporation cocktail was plated onto LB agar plates supplemented with ampicillin. The DNA sequence of the successful transformant was screened by restriction analysis, and confirmed via sequencing with an Applied Biosystems automated sequencer using fluorescent terminator dye tags at the UCSF Biomolecular Resource Center.

***Cloning: Construction of pACYC-F*:** Using the primers designed for the pGEM-F-LOS-12 vector a fresh PCR amplification of the *lsgF* ORF was performed, using the pGEMLOS-4 vector as a template, with the Easy-A high fidelity thermophilic polymerase cocktail (Stratagene). The PCR product was digested with Xho I while the pACYC184 vector was digested with Sal I. The digested PCR product and vector were then purified and concentrated to 6 $\mu$l using Zymo kits. Two $\mu$l of each were run on a gel to assess concentration and purity, then the Xho I digested PCR product was ligated into the Sal I site of pACYC184 vector (interrupting the Tet$^r$ ORF) using the Rapid DNA Ligation Kit (Roche Applied Research). The ligation was digested for one *hour* with Sal I to reduce background recircularized vector. It was then run through a Zymo kit and transformed into electrocompetent XL-1 Blue cells using a Bio-Rad MicroPulser electroporator. The electroporation cocktail was plated onto LB agar plates supplemented with chloramphenicol. The DNA sequence of the successful transformant was screened by restriction analysis, and confirmed via sequencing with an Applied Biosystems automated sequencer using fluorescent terminator dye tags at the UCSF Biomolecular Resource Center.

***Construction of pACYC-EF*:** Using the upstream primer designed for the pGEM-E-LOS-12 vector and the downstream primer designed for the pGEM-F-LOS-12 vector, a PCR amplification of the *lsgEF* ORFs was performed, using the pGEM-LOS-4 vector as a template, with the Easy-A high fidelity thermophilic polymerase cocktail (Stratagene). The PCR product was digested with Xho I while the pACYC184 vector was digested

with Sal I. The digested PCR product and vector were purified and concentrated to 6 $\mu$l using Zymo kits. Two $\mu$l of each were run on a gel to assess concentration and purity and then the Xho I digested PCR product was ligated into the Sal I site of pACYC184 vector (interrupting the $Tet^r$ ORF) using the Rapid DNA Ligation Kit (Roche Applied Research). The ligation was digested for one *hour* with Sal I to reduce background recircularized vector. It was then run through a Zymo kit and transformed into electrocompetent XL-1 Blue (Stratagene) cells using a Bio-Rad MicroPulser electroporator. The electroporation cocktail was plated onto LB agar plates supplemented with chloramphenicol. The DNA sequence of the successful transformant was screened by restriction analysis, and confirmed via sequencing with an Applied Biosystems automated sequencer using fluorescent terminator dye tags at the UCSF Biomolecular Resource Center.

***Cloning: Construction of Two-Vector Strains:*** For each of the two-vector strains in Table 3-4, *E. coli* containing the relevant pACYC184 vector (pACYC-F or pACYC-EF, respectively) were cultured overnight in 20 ml of LB supplemented with chloramphenicol. The following day, 5 ml of the overnight cultures were added to 60 ml LB:chloramphenicol and the bacteria was grown with shaking at 37°C until the bacteria reached an OD600 of 0.6. The bacteria were harvested by centrifugation at 5400 x *g* in a Jouan MR1812 centrifuge. The pellet was washed three times with 50 ml of ice-cold sterile deionized water and then resolubilized in 160 $\mu$l of ice-cold sterile deionized water. Forty $\mu$l aliquots of the now-electrocompetent pACYC vector containing bacteria

were transformed with their respective LOS-12 vectors using a Bio-Rad MicroPulser electroporator. The electroporation cocktail was plated onto LB agar plates supplemented with ampicillin and chloramphenicol. Qiagen mini-plasmid preps of the successful transformants were then screened via PCR for the presence of their respective vectors. PACYC184 has an origin of replication that is compatible with the Col E1 origin of the pGEM vectors.

***Table 3-4: Chimeric E.coli Transformants***

| Transformant | Vector 01 | Vector 02 | Relevant Hib ORFs |
|---|---|---|---|
| LOS-12 | pGEM-LOS-12 | | *lsgG* |
| CG | pGEM-C-LOS-12 | | *lsgC, lsgG* |
| DG | pGEM-D-LOS-12 | | *lsgD, lsgG* |
| EG | pGEM-E-LOS-12 | | *lsgE, lsgG* |
| FG | pGEM-F-LOS-12 | | *lsgF, lsgG* |
| CFG | pGEM-C-LOS-12 | pACYC-F | *lsgC, lsgF, lsgG* |
| DFG | pGEM-D-LOS-12 | pACYC-F | *lsgD, lsgF, lsgG* |
| EFG | pGEM-E-LOS-12 | pACYC-F | *lsgE, lsgF, lsgG* |
| CEFG | pGEM-C-LOS-12 | pACYC-EF | *lsgC, lsgE, lsgF, lsgG* |
| DEFG | pGEM-D-LOS-12 | pACYC-EF | *lsgD, lsgE, lsgF, lsgG* |
| LOS-05 | pGEM-LOS-5 | | *lsgC, lsgD, lsgE, lsgF, lsgG* |

***LPS Extraction:*** The LPS of all strains were prepared using the modified hot phenol extraction procedure [21,22]. An additional DNAse incubation was performed to reduce recombinant DNA contamination from the pGEM and pACYC vectors used to confer the LPS phenotypes. Briefly, bacteria was harvested by centrifugation and suspended in 2.0 ml of phosphate-buffered saline (PBS). The suspension was then centrifuged at 10,000 x *g* for 15 min at 4°C and the supernatant discarded. The pellet was washed once more with PBS and the resulting pellet was suspended in 600 $\mu$l of deionized water and

incubated in a 65°C water bath for 15 min. An equal volume of hot (65°C) 90% phenol was added and the mixture was incubated at 65°C for 35 min with frequent vortexing. The suspension was chilled on ice and then centrifuged at 8500 x *g* for 20 min, 4°C. The supernatant was transferred to a 15 ml conical tube, and the phenol fraction was re-extracted with 600 μl of deionized water. The aqueous phases were pooled, and 1/10 sample volume of 3 M sodium acetate pH 5.2 and three volumes of ethanol were added. The samples were then placed at –20°C overnight to precipitate the LPS. The precipitate was centrifuged at 10,000 x *g*, 4°C for 15 min and the supernatant discarded. The sample was resolubilized in 200 μl deionized water and re-precipitated. The precipitate was dried and resolubilized in 500 μl of 60 mM Tris-Cl (pH 8.0) + 2 mM $MgCl_2$. 2 μg/ml of DNAse (Roche Applied Research) was added and the samples were incubated at 37°C for 2 *hours*. The samples were then precipitated overnight as described above and the resulting pellets were re-solubilized in 1.5 ml of deionized water. The samples were then ultracentrifuged in a TL-100 ultracentrifuge (Beckman) at 120,000 x *g* for 2 *hours*. The supernatants were discarded and the contact lens pellets were re-solubilized in 200 μl of deionized water and evaporated to dryness in a Speed-Vac concentrator (Savant). The dried samples were stored at –20°C.

***O-Deacylation of LPS:*** To make LPS samples more amenable to mass spectrometric analysis, LPS samples were treated with anhydrous hydrazine to remove their O-linked lipid chains. Briefly: dried samples of LPS (~100 μg) were treated with 100 μl of anhydrous hydrazine (Sigma) and incubated at 37°C for 30 min in microcentrifuge tubes

with occasional vortexing [23]. After the reaction, the samples were placed in an ice bath and 500 $\mu$l of chilled acetone was added dropwise to precipitate the O-deacylated LPS. Samples were allowed to sit at –20°C for about 1 *hour* and then they were centrifuged at 12,000 x *g* at 4°C. The supernatants were removed and the O-deacylated LPS pellets were dissolved in 100 $\mu$l of deionized water and evaporated to dryness on a Speed-Vac concentrator.

***Preparation of Oligosaccharide (OS):*** Extracted LPS were hydrolyzed in 1% acetic acid (2 mg LPS/mL) for 2 *hour*s at 100°C with stirring to cleave the KDO-lipid A bond and release free OS. The hydrolysates were centrifuged at 5000 x *g* for 20 min at 4°C and the supernatants removed. Pellets were washed with 1 mL deionized water and centrifuged again. The supernatants and washings were pooled and dried in a Speed-Vac concentrator. OS samples were resolubilized in 50 $\mu$l of deionized water and loaded into a Dispo-Biodialyzer (Harvard Apparatus) with a 1000 MWCO membrane, and dialyzed against deionized water overnight. The samples were then collected from the dialyzer by centrifugation at 2000 x *g* and dried in a Speed-Vac concentrator. The dry samples were stored at –20°C [24].

***HF Treatment of O-LPS and OS:*** To remove phosphate and phosphoethanolamine residues, O-LPS or OS samples (~100 μg) were placed in 1.5 ml polypropylene tubes. The samples were treated with 25 μl cold 48% aqueous HF (Sigma). The samples were

then kept at 4°C for 16 *hours*. Aqueous HF was evaporated under a stream of $N_2$ in a propylene dessicator containing NaOH pellets. The dessicator was connected to the house vacuum via an in-line NaOH trap to remove HF [24].

***Sample Preparation for MALDI:*** To remove salts and other low-molecular-weight contaminants from O-LPS and HF-treated O-LPS (HF-O-LPS) the samples were drop dialyzed. Approximately 100 $\mu$g of dry sample was dissolved in 20 $\mu$l of deionized water. Then 10 $\mu$l aliquots were desalted by drop dialysis on VSWP 0.025 $\mu$m nitrocellulose membranes (Millipore) over deionized water for 2 *hours*. Samples were recovered from the membranes, evaporated to dryness, and redissolved in 10 $\mu$l of deionized water for mass spectrometric analysis.

***MALDI-TOF Mass Spectrometry:*** To determine molecular weights of HF-O-LPS samples, they were analyzed on a Voyager MALDI-TOF instrument (Applied Biosystems) equipped with a nitrogen laser (337 nm). All spectra were recorded in the negative-ion mode using delayed extraction conditions with an accelerating voltage of 20,000 V. Briefly, a 1 $\mu$l aliquot containing ~10 $\mu$g of a dialyzed HF-O-LPS sample was delivered into a 0.5 ml microcentrifuge tube containing a few beads of Dowex-50W-X8 resin (100-200 mesh, $NH_4^+$ form) and allowed to sit for 1 min. One $\mu$l of the matrix solution (a saturated solution of 2,5-dihydroxybenzoic acid in acetone) was then added to the sample and mixed briefly. Finally, the samples were spotted on a stainless steel

MALDI target and allowed to air-dry. Approximately 150 laser shots were recorded for each sample and averaged. The spectra were smoothed with a 19-point Savitsky-Golay function and mass-calibrated using an external mass calibrant consisting of renin substrate tridecapeptide, oxidized insulin chain B, and bovine insulin (Sigma). Also, a two-point internal calibration was performed to correct the external calibration using the monophosphoryl form of the defined *E. coli* core LPS deprotonated molecular ions with $(M-H)^-$ at *m/z* of 2375.31 and 2729.62 [1].

***ELISA Analysis*** *(Apicella lab at the University of Iowa):* ELISA studies were performed by a modified whole cell ELISA method [17]. Briefly, microtiter plate wells were coated with 100 μl of a suspension of organisms in distilled water. The organisms were dried onto the microtiter well surface in a dry 37°C incubator for 24 *hour*s. Prior to the ELISA the dried organisms were treated with 100 μl of 0.1 unit/mL neuraminidase in 5 mM NaOAc, 150 mM NaCl, and 4 mM$CaCl_2$, pH 5.6, for 2 *hour*s at 37°C. Control wells were not treated with neuraminidase. The wells were washed and the ELISA was performed. Two monoclonal antibodies were used in the ELISA studies. Monoclonal antibody 3F11 was reacted with each well at a dilution of 1:320 and monoclonal antibody 6B4 was studied at a dilution of 1:640. Secondary antibody (goat anti-mouse IgM alkaline phosphatase conjugate, Jackson ImmunoResearch) was reacted with each well at a dilution of 1:2000. P-nitrophenylphosphate was used to develop color at a concentration of 1.0 mg/ml. Results were read in a BioTek Model EL311 ELISA reader at 405 nm.

***Methylation Analysis:*** For the determination of glycosyl linkage positions, the samples were subjected to methylation analysis. First, purified OS was permethylated, hydrolyzed, reduced, and acetylated. The resultant partially methylated alditol acetates (PMAAs) were then analyzed by gas chromatography-mass spectrometry (GC-MS) as described by York *et al* [25]. The chemical derivatization procedures were performed by the Complex Carbohydrate Resource Center at The University of Georgia (CCRC, http://www.ccrc.uga.edu/). The derivatized samples were then subjected to GC-MS analysis in our laboratory at UCSF.

Initially, an aliquot of each sample was permethylated by the method of Ciukanu and Kerek [26] (treatment with sodium hydroxide and methyl iodide in dry DMSO). The permethylation was repeated twice in order to aid complete methylation of the polymer. Following sample workup, the permethylated material was hydrolyzed using 2 M trifluoroacetic acid (2 *hours* in sealed tube at 121°C), reduced with $NaBD_4$, and acetylated using acetic anhydride/trifluoroacetic acid.

Upon receipt of the PMAA samples from the CCRC, they were dried under $N_2$ and resuspended in 20 μL dichloromethane. One μL of each sample was used for injection. GC-MS analysis was on a Varian 2100T ion-trap GC/MS/MS with a 3900 GC (Varian Inc, Walnut Creek, CA) operating in splitless mode with a DB-1 capillary column (#122-1033, 30 m X 0.25 mm i.d., 100% methyl polysiloxane, 1 μm film thickness; Agilent, Inc., Palo Alto, CA). GC conditions: Injector was at 250°. Initial column temperature was 90°C for 4 min and then ramped at 20°C per min to 160°C and held for 2 min. Column was then ramped at 2° per min to 220° and then 10° per min to

250° and held for 11 min making a total run time of 53.5 min. MS conditions: Electron impact ionization, positive ion mode, scan time 0.63 seconds, and mass range 30 - 650 *m/z*.

## RESULTS

***Analysis HF-O-LPS from single gene + lsgG transformants:*** To screen these transformants for Hib ORF activity, aliquots of LPS from the CG, DG, EG, and FG transformants (Table 3-4) were treated with anhydrous hydrazine to remove O-linked fatty acids from the lipid A moiety. To further reduce heterogeneity the samples were treated with aqueous HF to remove phosphate and phosphoethanolamine. The negative ion MALDI-TOF spectra of the HF-treated O-deacylated, LPS samples (HF-O-LPS) are shown in Figure 3-1. The samples are somewhat heterogeneous, but the spectra clearly show that the HF-O-LPS of transformant FG is one hexose (+162 Da) larger than the HF-O-LPS of the other three transformants. The experimental molecular weight values as determined by MALDI-TOF (see Figure 3-1) were matched to calculated molecular weights (±1 Da) based on the composition of the *E. coli* core LPS and previous chimeric structures [1]. The calculated and experimental molecular weights of the observed species are detailed in Table 3-5 along with composition information for each mass. All four samples were found to exhibit moderate amounts of phosphate heterogeneity. All phosphoethanolamine residues were quantitatively removed by HF treatment, and each glycoform was typically observed as containing zero to three phosphates (a mixture of

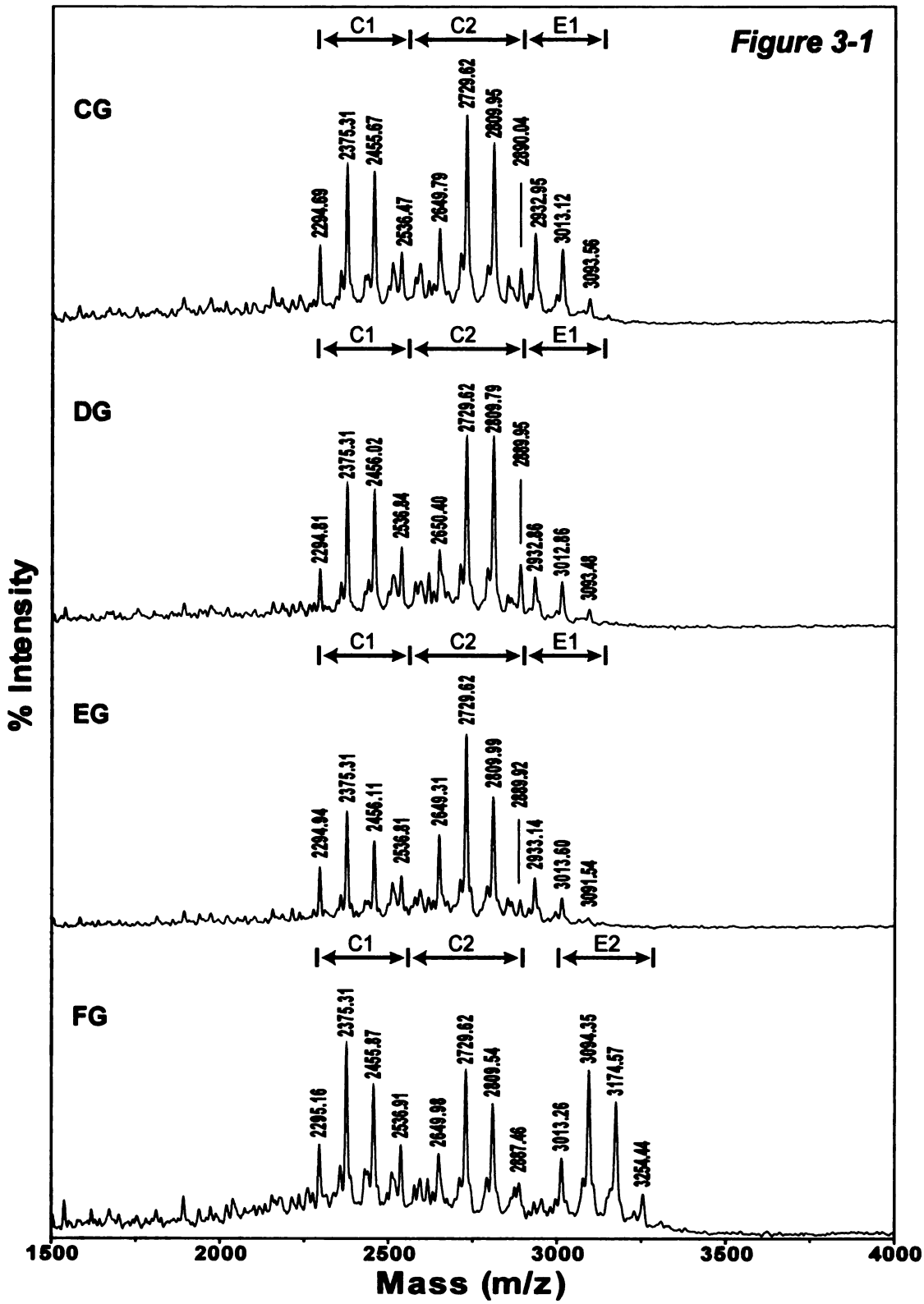



***Figure 3-1:*** Negative ion MALDI-TOF spectra of the O-deacylated, HF treated LPS from transformants CG, DG, EG, and FG. Each grouping of four peaks consists of the OLA, OLA+P, OLA+2P and OLA+3P versions of the same LPS species. Peaks are annotated in Table 3-5.

***Table 3-5: Molecular weights and proposed compositions of the HF treated, O-deacylated LPS from E. coli K-12 transformants* CG, DG, EG, FG, CFG, DFG, EFG, CEFG, DEFG, and LOS-05***

| | Experimental Average m/z (M-H)⁻ | | | | | | | | | | Calculated Average | | Proposed Compositions | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Transformant*: | CG | DG | EG | FG | CFG | DFG | EFG | CEFG | DEFG | LOS-05 | m/z (M-H)⁻ | | Hex | HexNAc | PO4 | Hep | Core‡ |
| Glycoform | | | | | | | | | | | | Glycoform | | | | | |
| E5 | | | | | | | | | | | 3823.45 | E5 | 3 | 3 | 3 | 1 | 1 |
| E5 | | | | | | | | | | 3744.53 | 3743.47 | E5 | 3 | 3 | 2 | 1 | 1 |
| E5 | | | | | | | | | | 3662.30 | 3663.49 | E5 | 3 | 3 | 1 | 1 | 1 |
| E5 | | | | | | | | | | 3582.72 | 3583.51 | E5 | 3 | 3 | | 1 | 1 |
| E4 | | | | | | | 3621.61 | 3620.56 | 3621.53 | 3620.81 | 3620.26 | E4 | 3 | 2 | 3 | 1 | 1 |
| E4 | | | | | | | 3540.37 | 3541.00 | 3540.92 | 3540.74 | 3540.28 | E4 | 3 | 2 | 2 | 1 | 1 |
| E4 | | | | | | | 3460.33 | 3460.08 | 3460.28 | 3460.30 | 3460.30 | E4 | 3 | 2 | 1 | 1 | 1 |
| E4 | | | | | | | 3379.13 | 3379.17 | 3380.27 | 3378.52 | 3380.32 | E4 | 3 | 2 | | 1 | 1 |
| E3 | | | | | | | | | | | 3458.12 | E3 | 2 | 2 | 3 | 1 | 1 |
| E3 | | | | | | | 3379.13 | 3379.17 | | 3378.52 | 3378.14 | E3 | 2 | 2 | 2 | 1 | 1 |
| E3 | | | | | | | 3298.32 | 3298.46 | | 3297.88 | 3298.16 | E3 | 2 | 2 | 1 | 1 | 1 |
| E3 | | | | | | | 3218.00 | 3217.66 | | | 3218.18 | E3 | 2 | 2 | | 1 | 1 |
| E2 | | | | 3254.44 | | | | | | | 3254.92 | E2 | 2 | 1 | 3 | 1 | 1 |
| E2 | | | | 3174.57 | 3175.17 | 3174.94 | | | | | 3174.94 | E2 | 2 | 1 | 2 | 1 | 1 |
| E2 | | | | 3094.35 | 3094.77 | 3094.84 | | | | | 3094.96 | E2 | 2 | 1 | 1 | 1 | 1 |
| E2 | | | | 3013.26 | 3014.54 | 3014.10 | | | | | 3014.98 | E2 | 2 | 1 | | 1 | 1 |
| E1 | 3093.56 | 3093.48 | 3091.54 | | | | | | | | 3092.78 | E1 | 1 | 1 | 3 | 1 | 1 |
| E1 | 3013.12 | 3012.86 | 3013.60 | | | | | | | | 3012.80 | E1 | 1 | 1 | 2 | 1 | 1 |
| E1 | 2932.95 | 2932.86 | 2933.14 | | | | | | | | 2932.82 | E1 | 1 | 1 | 1 | 1 | 1 |
| E1 | 2853.11 | 2852.49 | 2853.25 | | | | | | | | 2852.84 | E1 | 1 | 1 | | 1 | 1 |
| C2 | 2890.04 | 2889.95 | 2889.92 | 2887.46 | | | | | | | 2889.58 | C2 | 1 | | 3 | 1 | 1 |
| C2 | 2809.95 | 2809.79 | 2809.99 | 2809.54 | 2809.60 | 2809.52 | 2809.54 | 2810.14 | 2810.46 | 2810.06 | 2809.60 | C2 | 1 | | 2 | 1 | 1 |
| C2 | 2729.62 | 2729.62 | 2729.62 | 2729.62 | 2729.62 | 2729.62 | 2729.43 | 2729.62 | 2730.13 | 2729.62 | 2729.62 | C2 | 1 | | 1 | 1 | 1 |
| C2 | 2649.79 | 2650.40 | 2649.31 | 2649.98 | 2649.04 | 2649.22 | 2648.84 | 2653.55 | 2651.97 | 2649.69 | 2649.64 | C2 | 1 | | | 1 | 1 |
| C1 | 2536.47 | 2536.84 | 2536.81 | 2536.91 | 2536.52 | 2536.56 | 2537.34 | 2537.53 | 2537.18 | 2537.65 | 2535.27 | C1 | | | 3 | | 1 |
| C1 | 2455.67 | 2456.02 | 2456.11 | 2455.87 | 2455.82 | 2456.04 | 2456.35 | 2456.07 | 2456.16 | 2455.31 | 2455.29 | C1 | | | 2 | | 1 |
| C1 | 2375.31 | 2375.31 | 2375.31 | 2375.31 | 2375.31 | 2375.31 | 2375.34 | 2375.31 | 2375.29 | 2375.31 | 2375.31 | C1 | | | 1 | | 1 |
| C1 | 2294.69 | 2294.81 | 2294.94 | 2295.16 | 2294.98 | 2294.98 | 2294.44 | 2294.68 | 2294.76 | 2293.84 | 2295.33 | C1 | | | | | 1 |

*A list of the relevant ORFs expressed by each transformant is featured in Table 4. ‡Core = 3 Hex, 3 Hep, 2 KDO, 1 aphosphoryl O-deacyl Lipid A. The residue mass values (average mass) for LPS structural moities are as follows: Hex (162.142), Hep (192.169), HexNAc (203.195), KDO (220.179), phosphate (79.980), PEA (123.048), and aphosphoryl O-deacyl Lipid A (775.034).

the dephosphoryl, monophosphoryl, diphosphoryl, and occasionally triphosphoryl forms). Further HF treatment would have likely reduced the abundance of some of the more highly phosphorylated species, but would have degraded the acid-labile ketosidic linkage between the *E. coli* core LPS and the lipid-A moiety [27]. There is also a small amount of heterogeneity due to water loss, some of the larger peaks have a small –18 *m/z* anhydro peak at their base, a process our laboratory has reported on [1].

In all four MALDI-TOF spectra shown in Figure 3-1, the molecular ions in the mass range between *m/z* 2294 and 3093 represent the major *E. coli* core LPS structures. The peaks that comprise the un-extended *E. coli* LPS core [1] are labeled as the Core-1 (C1) glycoforms in Table 3-5. Also evident are the peaks that correspond to the extended *E. coli* LPS core [1] labeled as the Core-2 (C2) glycoforms in Table 3-5.

The first three spectra of Figure 3-1 correspond to the HF-O-LPS of the CG, DG, and EG transformants (Table 3-4). In addition to the C1 and C2 glycoforms, these spectra feature additional molecular ions in the mass range between *m/z* 2852 and 3093. These ions correspond to a single HexNAc extension (+ 203 Da) of the C2 glycoform and their masses are labeled as the extension-1 (E1) glycoforms in Table 3-5. This extension presumably occurs as a result of the presence of LsgG and a functional WecA, and has been previously documented [1]. All three of these spectra were identical to a control spectrum of HF-O-LPS from *E. coli* K-12 transformed with the background pGEM-LOS-12 vector (not shown). From these three spectra we observe that the LsgC, LsgD, and LsgE gene products were incapable of recognizing the E1 glycoform as a substrate.

The fourth MALDI-TOF spectrum of Figure 3-1 corresponds to the HF-O-LPS from transformant FG. This spectrum also features the C1 and C2 glycoforms, but the E1

glycoforms are not seen. Instead, there is a group of molecular ions between *m/z* 3014 and 3254. This latter set of peaks corresponds to a single Hex extension of the E1 glycoform. The presence of LsgF leads to the extension the E1 glycoform by a single Hex (+162 Da). All calculated and experimental masses as determined by MALDI-TOF analyses are shown in Table 3-5 and are reported as average masses.

***Analysis HF-O-LPS from two gene + lsgG transformants:*** Aliquots of LPS from the CFG, DFG and EFG transformants (Table 3-4) were also treated with anhydrous hydrazine to remove O-linked fatty acids from the lipid A moiety and were partially dephosphorylated by treatment with aqueous HF. The negative-ion MALDI-TOF spectra of these three HF-O-LPS samples are shown in Figure 3-2. All three HF-O-LPS spectra in Figure 3-2 feature the C1 and C2 glycoforms. The HF-O-LPS spectra of transformant CFG and transformant DFG also feature the E2 glycoform, but no further extensions beyond the glycoform synthesized by LsgF and LsgG alone. The third spectrum in Figure 3-2 corresponds to the HF-O-LPS of transformant EFG and shows a dramatic increase in mass relative to the peaks seen in other two spectra. The set of peaks between *m/z* 3218 and 3458 correspond to the addition of HexNAc (+203 Da) to glycoform E2 and their masses are labeled as the E3 glycoforms in Table 3-5. The peak set between *m/z* 3380 and 3620 corresponds to the addition of Hex (+162 Da) to glycoform E3 and their masses are labeled as the E4 glycoforms in Table 3-5. There is some overlap between the *m/z* 3378.1 peak of glycoform E3 and the *m/z* 3380.3 peak of glycoform E4. Similarly *m/z* 3458.12 of E3 and *m/z* 3460.30 of glycoform E4 overlap as well. While a

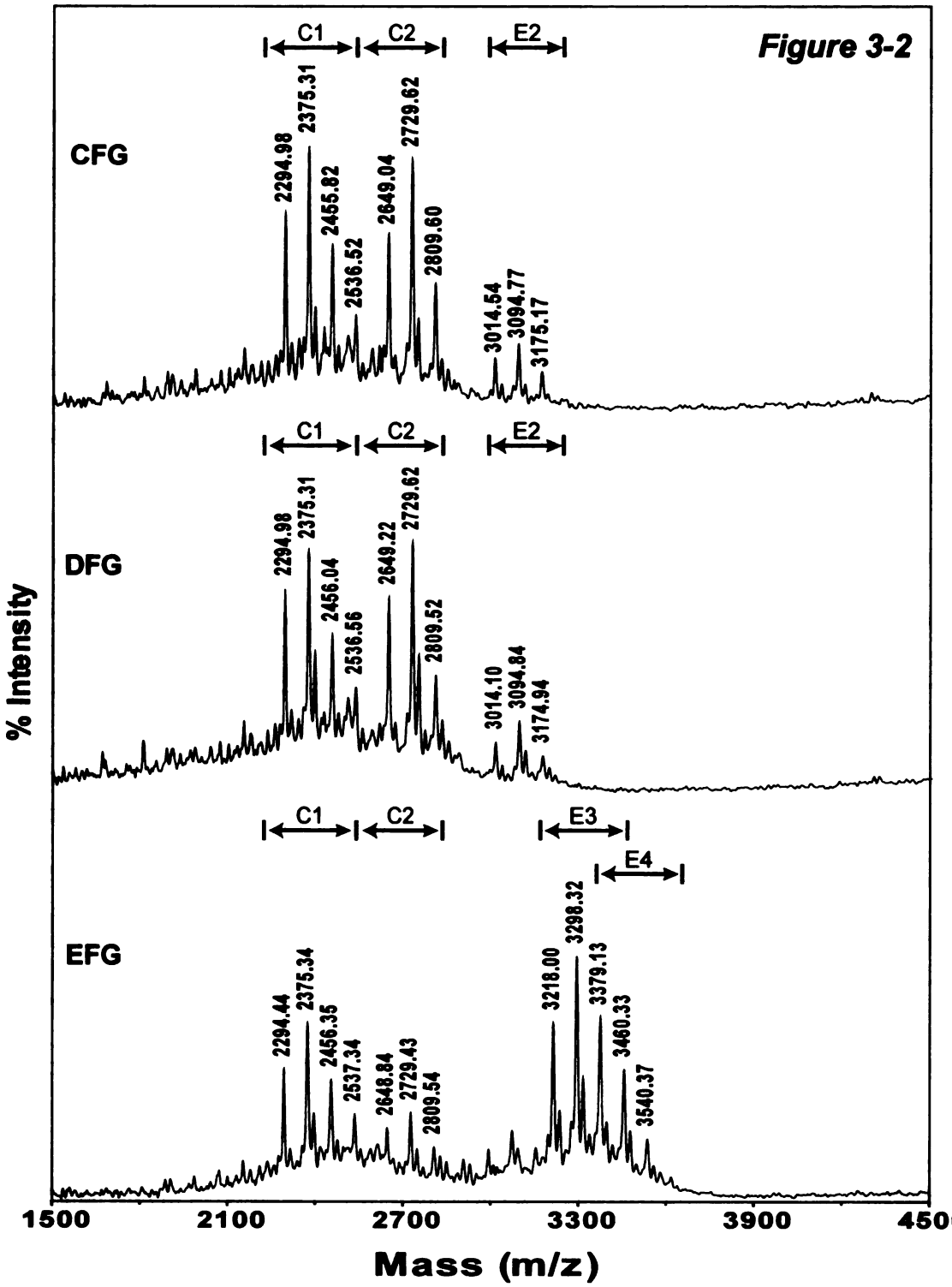



***Figure 3-2:*** Negative ion MALDI-TOF spectra of the O-deacylated, HF treated LPS from transformants CFG, DFG, and EFG. Each grouping of four peaks consists of the OLA, OLA+P, OLA+2P and OLA+3P versions of the same LPS species. Peaks are annotated in Table 3-5.

higher resolution instrument could have resolved these peaks, we have found TOF-MS instruments in linear mode with shorter flight times are more effective for observing O-LPS and HF-O-LPS than higher mass accuracy mass spectrometers with reflectron optics. This is likely the result of extensive metastable fragmentation in the field-free regions [27]. The appearance of the E3 glycoform in the spectrum of the HF-O-LPS of transformant EFG suggests that the presence of LsgE leads to the HexNAc extension glycoform E2. The appearance of glycoform E4 in this spectrum implies that LsgF is also capable recognizing glycoform E3 as a substrate.

***Analysis HF-O-LPS from three gene + lsgG transformants:*** Aliquots of LPS from the CEFG, DEFG and LOS-05 transformants (Table 3-4) were also treated with anhydrous hydrazine to remove O-linked fatty acids from the lipid A moiety, and partially dephosphorylated via treatment with aqueous HF. The negative ion MALDI-TOF spectra of these HF-O-LPS samples are shown in Figure 3-3. The first spectrum corresponds to the HF-O-LPS of the CEFG transformant and displays glycoforms C1, C2, E3, and E4 in similar abundance to that of the HF-O-LPS of transformant EFG from Figure 3-2. The presence of LsgC does not seem to have any effect on the observed LPS glycoforms. The second spectrum of Figure 3-3 corresponds to the HF-O-LPS from transformant DEFG. This spectrum displays a similar array of peaks with one key difference: the peaks corresponding to glycoform E3 are largely absent from the DEFG HF-O-LPS spectrum. This implies that the presence of LsgD also results in a Hex extension of glycoform E3, and that LsgD competes effectively with LsgF to extend glycoform E3. The fourth spectrum is of the HF-O-LPS from the LOS-05 transformant (Table 3-4) and displays a

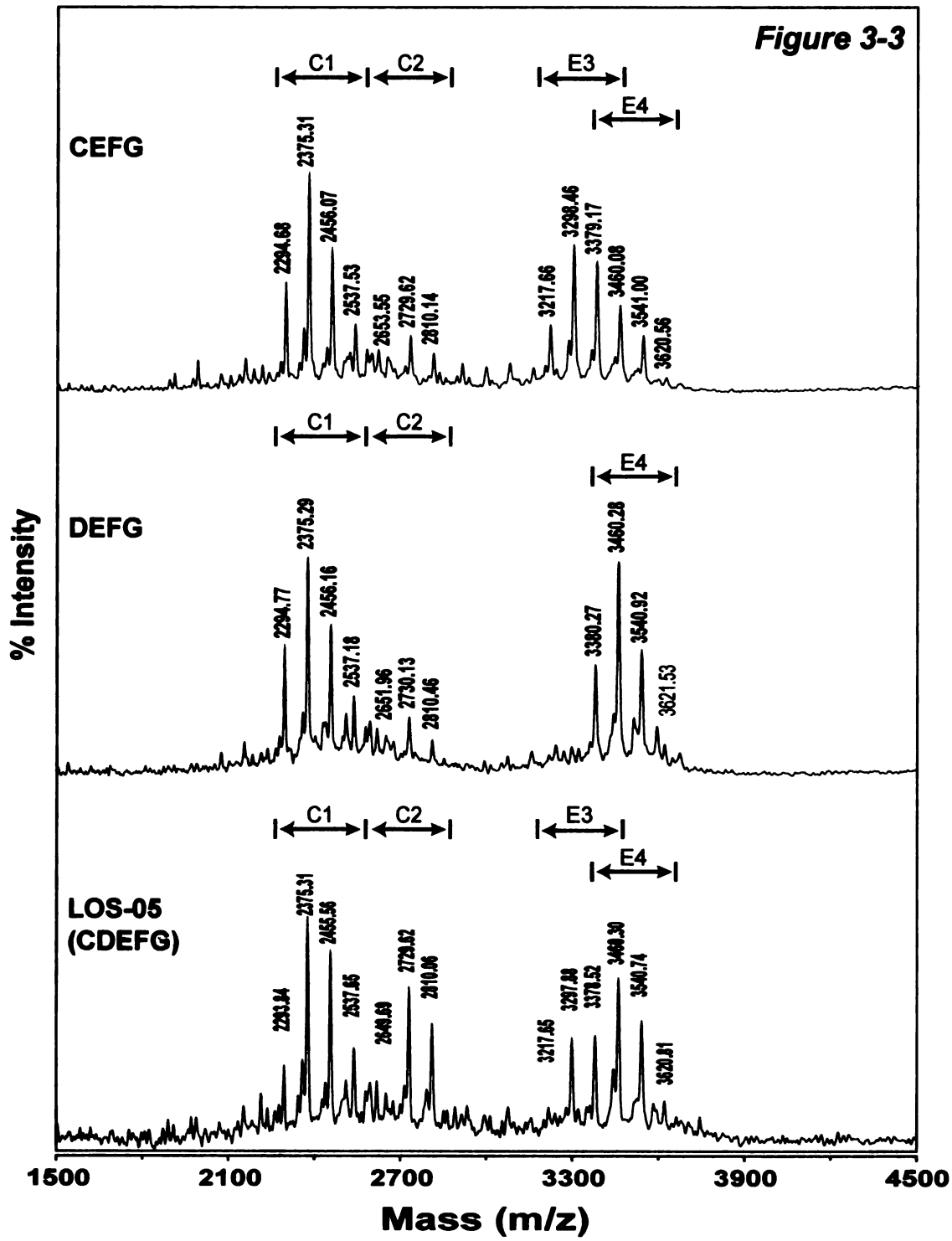



***Figure 3-3:*** Negative ion MALDI-TOF spectra of the O-deacylated, HF treated LPS from transformants CEFG, DEFG, and LOS-05. Each grouping of four peaks consists of the OLA, OLA+P, OLA+2P and OLA+3P versions of the same LPS species.
Peaks are annotated in Table 3-5.

similar pattern to the above spectra, with one key difference: there is a set of small peaks to the right of the E4 glycoform that may be a HexNAc addition. However, several attempts to clean up these HF-O-LPS samples with drop dialysis did not result in MALDI-TOF spectra with resolution of these peaks suitable to verify this addition. Two peaks corresponding to glycoform E3 appear in this spectrum as well. This implies that LsgC interferes with the action of LsgD or LsgF in some way although this precise mechanism is not known.

***ELISA Analysis of the EFG, CEFG, DEFG and LOS-05 Transformants:*** Plates of intact organisms from each of the transformants capable of synthesizing glycoform E4 were sent to our collaborators at the University of Iowa for ELISA analysis (Dr. Michael Apicella, Department of Microbiology). As shown in Table 3-6, transformants DEFG and LOS-05 react strongly with both MAbs 3F11 and 6B4.

***Table 3-6: ELISA of Mab 3F11 and Mab 6B4 LPS Epitopes on E4 Glycoform containing transformants with and without Neuraminidase treatment***

| | Neurimidase | 3F11 | 6B4 |
|---|---|---|---|
| **EFG** | - | 0.000[a] | 0.000[a] |
| | + | 0.000[a] | 0.000[a] |
| **CEFG** | - | 0.000[a] | 0.000[a] |
| | + | 0.000[a] | 0.000[a] |
| **DEFG** | - | 0.677[b] | 0.767[c] |
| | + | 0.637[b] | 0.843[c] |
| **CDEFG** | - | 0.557[b] | 0.781[c] |
| | + | 0.567[b] | 0.841[c] |

a: values were less than the detection limit at dilutions as low as 1:40 for both antibodies.
b: Mab dilution of 1:320 c: Mab dilution of 1:640

Monoclonal antibody 3F11 has been shown to recognize terminal N-acetyl lactosamine (Galβ1–4 GlcNAc) [1,17,28]. Monoclonal antibody 6B4 has also been shown to recognize terminal N-acetyl lactosamine [29]. Neither transformant EFG nor CEFG were shown to react with either antibody. Although the HF-O-LPS from transformants EFG and CEFG had mass profiles similar to those of transformants DEFG and LOS-05 via MALDI, this ELISA data suggests that they are structurally different with regards to stereochemistry or linkage.

***Linkage Analysis of the EFG, CEFG, DEFG, and LOS-5 Transformants:*** HF-treated oligosaccharide fractions (HF-OS) from each of the transformants capable of synthesizing glycoform E4 were sent to the Complex Carbohydrate Research Center (University of Georgia) for conversion to partially methylated alditol acetates (PMAAs) for linkage analysis. The derivatized PMAAs were analyzed via GCMS in our laboratory and the results of this analysis are featured in Table 3-7. These results are similar for the OS of all four transformants with one notable difference: there were negligible amounts of the PMAA corresponding to 1–4 GlcNAc present in the PMAA samples from transformants EFG and CEFG. It was, however, detected in the PMAA samples from transformants DEFG and LOS-05, which contain LsgD. This suggests that LsgF, which adds Gal at the 3-position of the terminal GlcNAc of Glycoform E1 [1], is also capable of adding a Gal to the 3-position of the terminal GlcNAc of Glycoform E3. The presence of the 1–4 GlcNAc extensions in LsgD containing transformants suggests that LsgD is a Gal1–4 transferase capable of competing effectively with LsgF for the E3 glycoform.

***Table 3-7: Methylation Analysis of the dephosphorlyated oligosaccharide fractions***
Peak areas were measured from the GC/MS total ion chromatograms, and values were normalized to the 3,6-Glc residue.

| Peak | EFG | CEFG | DEFG | LOS-05 |
|---|---|---|---|---|
| **T–Gal** | 1.1 | 0.9 | 1.3 | 1.5 |
| **2–Glc** | 0.9 | 0.8 | 0.9 | 1.0 |
| **3–Gal** | 0.6 | 0.3 | 0.6 | 0.6 |
| **6–Glc** | 0.7 | 0.6 | 0.7 | 1.0 |
| **3,6–Glc** | 1.0 | 1.0 | 1.0 | 1.0 |
| **T–Hep** | 0.4 | 0.6 | 0.6 | 0.5 |
| **3–Hep** | 0.2 | 0.1 | 0.3 | 0.2 |
| **7–Hep** | 0.5 | 1.0 | 0.6 | 0.6 |
| **3,7–Hep** | 0.2 | 0.1 | 0.3 | 0.2 |
| **T–GlcNac** | 0.5 | 0.3 | 0.1 | 0.2 |
| **4–GlcNAc** | | | 0.4 | 0.3 |
| **3–GlcNAc** | 0.4 | 0.2 | 0.2 | 0.4 |

## DISCUSSION

This study provides structural data to support our original hypothesis that stepwise addition of the *lsg* locus glycosyltransferase genes will support elucidation of the substrate specificity of each gene by determining the step of the synthesis of the LOS-05 glycoform in which it functions. In this manner the structure previously determined for the chimeric LPS from *E. coli* transformed with the LOS-05 vector [1] can be used as a template for deciphering the mass, composition and linkage data from the current study.

LOS-05 Structure: Gal1–4 GlcNAc1–3 Gal1–3 GlcNAc→CORE

The first step in the synthesis of the LOS-05 chimeric LPS is the addition of GlcNAc to the 7-position of the terminal L-glycero-D-manno-heptose from the *E. coli* core (glycoform E1). Phillips et al. [1] previously reported that this crucial step was

dependent on the presence of *lsgG* and the *E. coli* strain having a functional *wecA* gene. WecA mediates the transfer of UDP-GlcNAc to undecaprenyl phosphate with the formation of a phosphodiester bond [30,31]. *E. coli* typically employs this reaction to initiate the synthesis of O-antigen repeats in LPS biosynthesis. WecA transfers the GlcNAc to the undecaprenyl carrier which associates the nascent oligosaccharide chain to the inner membrane of the cytoplasm as the GlcNAc accepts sequentially transferred monosaccharides one unit at a time. Then it is exported to the periplasm and ligated to the core LPS (which is synthesized by another pathway) [1,30,31]. In the case of *E. coli* K-12, a rough strain that does not display an O-antigen on its LPS, it is our hypothesis that LsgG, which has homology to a transcriptional regulator, [1] is involved in the regulation of *wecA*, enhancing the production of chimeric LPS.

The second step in the synthesis of the LOS-05 chimeric LPS is the addition of galactose to the 3-position of the terminal GlcNAc of glycoform E1. Transformant FG was the only one gene + *lsgG* transformant in which this LPS structure was extended by one galactose. Given this data, one can therefore conclude that *lsgF* encodes a 1–3 galactosyltransferase, which is capable of recognizing GlcNAc on the pyrophosphoryl-undecaprenyl carrier lipid as a substrate.

The third step in the formation of the LOS-05 chimeric LPS is the addition of N-acetyl-glucosamine to the 3 position of the terminal Gal of glycoform E2. Transformant EFG was the only two gene + *lsgG* transformant in which this LPS structure was extended. Instead of a single sugar extension, however, we see a modular pattern in the mass spectrum that represents the addition of a GlcNAc as well as the addition of a GlcNAc-Gal extension of glycoform E2. The fact that no extension of the E2 glycoform

occurs in the CFG and DFG transformants, and that the two-sugar extension does not occur without LsgE present, implicates LsgE as the 1–3 N-acetyl glucosaminyl transferase which can recognize the terminal Gal1–3 GlcNAc moiety of the E2 glycoform as a substrate. It is possible that LsgF, or some endogenous *E. coli* transferase is capable of extending this GlcNAc by one Gal as well. The results of the linkage analysis, however, only show 3-GlcNAc, and no other di-linked GlcNAc. This linkage analysis data, coupled with the fact that GlcNAc-pyrophospate-undecaprenol is most likely not the endogenous substrate of LsgF [14,24], implicates LsgF as a Gal-transferase capable of recognizing GlcNAc1–3 Gal as well as GlcNAc-pyrophospate-undecaprenol as a substrate.

The fourth step in the synthesis of the LOS-05 glycoform is the addition of the above-mentioned galactose to the 3- or 4-position of the terminal GlcNAc of the E3 glycoform. The HF-O-LPS spectrum from transformant CEFG displays a similar pattern to that of the HF-O-LPS spectrum from transformant EFG. In both spectra HF-O-LPS from transformants EFG and CEFG display an abundance of glycoform E3 ions that is much higher relative to the abundance of glycoform E4 ions. The HF-O-LPS spectrum from transformant DEFG, however, displays a much higher abundance of glycoform E4 ions relative to glycoform E3 ions. In fact the glycoform E3 ions are nearly nonexistent. The spectrum of HF-O-LPS from the LOS-05 transformant also displays a more prominent abundance of glycoform E4 ions relative to glycoform E3 ions, but the difference is not as pronounced as it is in the spectrum of transformant DEFG. The methylation analysis results (Table 3-7) confirm the presence of 4-GlcNAc in the DEFG and LOS-05 samples. The ELISA analysis (Table 3-6) also indicates that Mab 3F11 and

Mab 6B4 (which recognize the terminal Galβ1–4 GlcNAc epitope) only bind to organisms from transformants DEFG and LOS-05. The sum of these data together implicate LsgD as a 1–4 galactosyltransferase that is capable of recognizing GlcNAc1–3-Gal as a substrate.

## CONCLUSIONS

We have determined the synthetic properties of four glycosyltransferases, LsgC, LsgD, LsgE and LsgF, and have assigned functions to all but LsgC. The functional assignment of these glycosyltransferases has previously eluded assignment [1,9,11,14,16-18]. A brief summary of my conclusions regarding the synthetic properties of these enzymes is shown in Table 3-8.

***Table 3-8: Specificity and Selectivity Conclusions***

| Gene Product | Monomer Added | Linkage added | Nascent Chain Disaccharides Extended | Nascent Chain Disaccharides not Extended |
|---|---|---|---|---|
| LsgF | Galactose | 1–3 | GlcNAc1–3 Gal<br>GlcNac1– $PO_4$-undecaprenol | Gal1–4 GlcNAc<br>Gal1–3 GlcNAc |
| LsgE | N-Acetyl Glucosamine | 1–3 | Gal1–3GlcNAc | Gal1–4 GlcNAc<br>GlcNAc1–3 Gal<br>GlcNac1– $PO_4$-undecaprenol |
| LsgD | Galactose | 1–4 | GlcNAc1–3Gal | Gal1–4 GlcNAc<br>Gal1–3 GlcNAc<br>GlcNac1– $PO_4$-undecaprenol |
| LsgC | ND | ND | | Gal1–4 GlcNAc<br>GlcNAc1–3 Gal<br>Gal1–3 GlcNAc<br>GlcNac1– $PO_4$-undecaprenol |

Knowledge of the synthetic properties of the Lsg gene products will assist in the elucidation of the molecular mechanisms of Hib LOS synthesis and its contribution to pathogenicity. Additionally, these genes have been successfully cloned and have been used to express oligosaccharide. These glycosyltransferases can be put to use as part of an enzymatic or chemoenzymatic synthesis of complex carbohydrate [4]. *E. coli* K-12 transfected with vectors from this study could be incorporated into a bacterial metabolic pathway engineering method for the large-scale synthesis of complex carbohydrates [6,32].

Furthermore, we have synthesized the lacto-N-neotetraose (Gal1–4-GlcNAc1–3Gal1–3GlcNAc) [33] glycoform in this *E. coli* system, which is typically associated with *Neisseria gonorrhoeae* and *Neisseria meningitidis* lipooligosaccharide (LOS) [10,33]. This tetrasaccharide has been implicated in the invasion of human urethral epithelial cells [33,34]. Large-scale preparation of LPS from transformant DEFG could potentially be a safer and more reliable way to obtain large quantities of this glycolipid. *N. gonorrhoeae* and *N. meningitidis* are both human pathogens [12] whereas transformant DEFG is not. Furthermore, *N. gonorrhoeae* and *N. meningitidis* LOS have been shown to demonstrate a high degree of phase variation [12], whereas transformant DEFG does not.

This multiplasmid expression system may potentially be expanded as a general method for the functional assignment of microbial glycosyltransferases. A small library of glycosyltransferase gene pGEM vectors that synthesize known disaccharide-acceptors in *E. coli* K-12 could be constructed. The ORF of a glycosyltransferase gene of unknown function would be cloned into pACYC184 and transfected into *E. coli*. These *E. coli*

could then be transfected with each pGEM vector from the disaccharide-acceptor library and screened for phenotypic variation. This method of functional assignment could potentially be easier than performing knockout and rescue studies in the endogenous organism, which may phase-vary its carbohydrate epitopes, or have multiple levels of degeneracy at each stage of carbohydrate synthesis.

As this chimeric synthesis methodology continues to be optimized and expanded to include more glycosyltransferase genes, it may well have its greatest potential as a general method for the production of specific oligosaccharides and lipid-linked oligosaccharides. One particularly exciting application of this technology might be the construction of carbohydrate-based vaccines, where carbohydrate epitopes from pathogenic bacteria could be incorporated into non-lethal strains. These strains could comprise the vaccine itself. Alternatively, there have been advances in the field of vaccine adjuvants based on the structure of Lipid A, such as monophosphoryl Lipid-A [35]. Biosynthetic chimeric LPS representing a carbohydrate vaccine of interest could possibly be dephosphorylated and used as a vaccine without the need for additional adjuvants.

Another potential application of dephosphorylated biosynthetic LPS would be as targeting ligands for tissue-specific drug delivery. Biosynthetic chimeric LPS representing a carbohydrate known to target tissue specific glycan receptors could be de-phosphorylated and incorporated into liposomes carrying a drug of interest. The liposome would potentially circulate in the bloodstream until its carbohydrate moieties encountered glycan receptors that would bind it and perhaps trigger internalization [36].

## REFERENCES


1. Phillips NJ, Miller TJ, Engstrom JJ, Melaugh W, McLaughlin R, Apicella MA, Gibson BW: **Characterization of chimeric lipopolysaccharides from Escherichia coli strain JM109 transformed with lipooligosaccharide synthesis genes (*lsg*) from Haemophilus influenzae.** *J Biol Chem* 2000, **275**:4747-4758.
2. Varki A: **Biological roles of oligosaccharides: all of the theories are correct.** *Glycobiology* 1993, **3**:97-130.
3. Koeller KM, Wong CH: **Complex carbohydrate synthesis tools for glycobiologists: enzyme-based approach and programmable one-pot strategies.** *Glycobiology* 2000, **10**:1157-1169.
4. Sears P, Wong CH: **Toward automated synthesis of oligosaccharides and glycoproteins.** *Science* 2001, **291**:2344-2350.
5. Hanson S, Best M, Bryan MC, Wong CH: **Chemoenzymatic synthesis of oligosaccharides and glycoproteins.** *Trends Biochem Sci* 2004, **29**:656-663.
6. Fort S, Birikaki L, Dubois MP, Antoine T, Samain E, Driguez H: **Biosynthesis of conjugatable saccharidic moieties of GM(2) and GM(3) gangliosides by engineered E-coli.** *Chemical Communications* 2005:2558-2560.
7. Preston A, Mandrell RE, Gibson BW, Apicella MA: **The lipooligosaccharides of pathogenic gram-negative bacteria.** *Crit Rev Microbiol* 1996, **22**:139-180.
8. Mandrell RE, McLaughlin R, Aba Kwaik Y, Lesse A, Yamasaki R, Gibson B, Spinola SM, Apicella MA: **Lipooligosaccharides (LOS) of some Haemophilus species mimic human glycosphingolipids, and some LOS are sialylated.** *Infect Immun* 1992, **60**:1322-1328.
9. Phillips NJ, Apicella MA, Griffiss JM, Gibson BW: **Structural studies of the lipooligosaccharides from Haemophilus influenzae type b strain A2.** *Biochemistry* 1993, **32**:2003-2012.
10. Harvey HA, Swords WE, Apicella MA: **The mimicry of human glycolipids and glycosphingolipids by the lipooligosaccharides of pathogenic neisseria and haemophilus.** *J Autoimmun* 2001, **16**:257-262.
11. Swords WE, Jones PA, Apicella MA: **The lipo-oligosaccharides of Haemophilus influenzae: an interesting array of characters.** *Journal of Endotoxin Research* 2003, **9**:131-144.
12. Phillips NJ, John CM, Reinders LG, Gibson BW, Apicella MA, Griffiss JM: **Structural models for the cell surface lipooligosaccharides of Neisseria gonorrhoeae and Haemophilus influenzae.** *Biomed Environ Mass Spectrom* 1990, **19**:731-745.
13. Gaucher SP, Cancilla MT, Phillips NJ, Gibson BW, Leary JA: **Mass spectral characterization of lipooligosaccharides from Haemophilus influenzae 2019.** *Biochemistry* 2000, **39**:12406-12414.
14. Spinola SM, Kwaik YA, Lesse AJ, Campagnari AA, Apicella MA: **Cloning and expression in Escherichia coli of a Haemophilus influenzae type b lipooligosaccharide synthesis gene(s) that encodes a 2-keto-3-deoxyoctulosonic acid epitope.** *Infect Immun* 1990, **58**:1558-1564.

15. Abu Kwaik Y, McLaughlin RE, Apicella MA, Spinola SM: **Analysis of Haemophilus influenzae type b lipooligosaccharide-synthesis genes that assemble or expose a 2-keto-3-deoxyoctulosonic acid epitope.** *Mol Microbiol* 1991, **5**:2475-2480.
16. Hood DW, Randle G, Cox AD, Makepeace K, Li J, Schweda EKH, Richards JC, Moxon ER: **Biosynthesis of Cryptic Lipopolysaccharide Glycoforms in Haemophilus influenzae Involves a Mechanism Similar to That Required for O-Antigen Synthesis.** *J. Bacteriol.* 2004, **186**:7429-7439.
17. Phillips NJ, McLaughlin R, Miller TJ, Apicella MA, Gibson BW: **Characterization of two transposon mutants from Haemophilus influenzae type b with altered lipooligosaccharide biosynthesis.** *Biochemistry* 1996, **35**:5937-5947.
18. McLaughlin RE, Lee N-G, Abu Kwaik Y, Spinola SM, Apicella MA: **Characterization and sequence analysis of the *lsg* (LOS synthesis genes) locus from *Haemophilus influenzae* type b.** *J. Endotoxin Res.* 1994, **1**:165-174.
19. Sambrook J, Russell DW: *Molecular cloning : a laboratory manual* edn 3rd. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press; 2001.
20. Stull B: **Personal Correspondence and His Own Molecular Biology Protocols.** Edited by: UCSF; 2004.
21. Inzana TJ: **Electrophoretic heterogeneity and interstrain variation of the lipopolysaccharide of Haemophilus influenzae.** *J Infect Dis* 1983, **148**:492-499.
22. Apicella MA, Griffiss JM, Schneider H: **Isolation and characterization of lipopolysaccharides, lipooligosaccharides, and lipid A.** *Methods Enzymol* 1994, **235**:242-252.
23. Helander IM, Lindner B, Brade H, Altmann K, Lindberg AA, Rietschel ET, Zahringer U: **Chemical structure of the lipopolysaccharide of Haemophilus influenzae strain I-69 Rd-/b+. Description of a novel deep-rough chemotype.** *Eur J Biochem* 1988, **177**:483-492.
24. Phillips NJ, Apicella MA, Griffiss JM, Gibson BW: **Structural characterization of the cell surface lipooligosaccharides from a nontypable strain of Haemophilus influenzae.** *Biochemistry* 1992, **31**:4515-4526.
25. York WS, Darvill AG, McNeil M, Stevenson TT, Albersheim P, Arthur Weissbach and Herbert W: **Isolation and characterization of plant cell walls and cell wall components.** In *Methods in Enzymology*. Edited by: Academic Press; 1986:3-40. vol Volume 118.]
26. Ciucanu I, Kerek F: **A simple and rapid method for the permethylation of carbohydrates.** *Carbohydrate Research* 1984, **131**:209-217.
27. Gibson BW, Engstrom JJ, John CM, Hines W, Falick AM: **Characterization of bacterial lipooligosaccharides by delayed extraction matrix-assisted laser desorption ionization time-of-flight mass spectrometry.** *Journal of the American Society for Mass Spectrometry* 1997, **8**:645-658.
28. Yamasaki R, Nasholds W, Schneider H, Apicella MA: **Epitope expression and partial structural characterization of F62 lipooligosaccharide (LOS) of Neisseria gonorrhoeae: IgM monoclonal antibodies (3F11 and 1-1-M) recognize non-reducing termini of the LOS components.** *Mol Immunol* 1991, **28**:1233-1242.

29. Apicella MA, Mandrell RE, Shero M, Wilson ME, Griffiss JM, Brooks GF, Lammel C, Breen JF, Rice PA: **Modification by sialic acid of Neisseria gonorrhoeae lipooligosaccharide epitope expression in human urethral exudates: an immunoelectron microscopic analysis.** *J Infect Dis* 1990, **162**:506-512.
30. Amer AO, Valvano MA: **Conserved aspartic acids are essential for the enzymic activity of the WecA protein initiating the biosynthesis of O-specific lipopolysaccharide and enterobacterial common antigen in Escherichia coli.** *Microbiology* 2002, **148**:571-582.
31. Alexander DC, Valvano MA: **Role of the rfe gene in the biosynthesis of the Escherichia coli O7-specific lipopolysaccharide and other O-specific polysaccharides containing N-acetylglucosamine.** *J Bacteriol* 1994, **176**:7079-7084.
32. Antoine T, Priem B, Heyraud A, Greffe L, Gilbert M, Wakarchuk WW, Lam JS, Samain E: **Large-scale in vivo synthesis of the carbohydrate moieties of gangliosides GM1 and GM2 by metabolically engineered Escherichia coli.** *Chembiochem* 2003, **4**:406-412.
33. Harvey HA, Jennings MP, Campbell CA, Williams R, Apicella MA: **Receptor-mediated endocytosis of Neisseria gonorrhoeae into primary human urethral epithelial cells: the role of the asialoglycoprotein receptor.** *Mol Microbiol* 2001, **42**:659-672.
34. Harvey HA, Porat N, Campbell CA, Jennings M, Gibson BW, Phillips NJ, Apicella MA, Blake MS: **Gonococcal lipooligosaccharide is a ligand for the asialoglycoprotein receptor on human sperm.** *Mol Microbiol* 2000, **36**:1059-1070.
35. Persing DH, Coler RN, Lacy MJ, Johnson DA, Baldridge JR, Hershberg RM, Reed SG: **Taking toll: lipid A mimetics as adjuvants and immunomodulators.** *Trends Microbiol* 2002, **10**:S32-37.
36. Eliaz RE, Szoka FC, Jr.: **Liposome-encapsulated doxorubicin targeted to CD44: a strategy to kill CD44-overexpressing tumor cells.** *Cancer Res* 2001, **61**:2592-2601.

# CHAPTER 4.

# FUNCTIONAL DETERMINATION OF THE *Haemophilus influenzae* SIALYLTRANSFERASE *lsgB* VIA EXPRESSION IN *Escherichia coli* K-12

## ABSTRACT

An *in vivo* biosynthesis strategy for the production of sialylated lipopolysaccharide (LPS) was adapted using the CMP-neuraminic acid (CMP-NeuAc) synthetase gene, *siaB*, from non-typeable *Haemophilus influenzae* (NTHI) [1] in a chimeric *E. coli* K-12 expression system. This system was used to evaluate the hypothesis that the glycosyltransferase gene *lsgB*, from *Haemophilus influenzae* type b (Hib), is indeed a sialyltransferase. Previously, this chimeric expression system was used to profile the synthetic properties (specificity and selectivity) of a number of glycosyltransferases from the *H. influenzae lsg* gene locus [2,3]. However, in all the backgrounds examined, we were unable to synthesize sialylated glycoforms, despite the expression of *lsgB*, a putative sialyltransferase [4]. In order to test the hypothesis that the obstacle to expression of sialylated glycoforms in *E. coli* K-12 was the lack of CMP-NeuAc, a vector containing a functional *SiaB* ORF (pJCSiaB) was transfected alongside pGEM-LOS-04 (which contains the entire *lsg* gene locus including *lsgB*). The LPS from this transformant and the LPS from *E. coli* K-12 transfected with pGEM-LOS-04 alone were subjected to MALDI-TOF MS analysis. The LPS from transformant pGEM-LOS-04 + pJCSiaB was identical to that of pGEM-LOS-04 with one distinct difference: the LPS from pGEM-LOS-04 + pJCSiaB contained one additional NeuAc, and one less 3-deoxy-D-manno-oct-2-ulosonic acid (KDO) than that of transformant pGEM-LOS-04.

## INTRODUCTION

In mammalian and avian systems, sialic acid (or N-acetylneuraminic acid, NeuAc) typically occupies the terminal position on oligosaccharide chains of cell-surface or serum glycoconjugates [5,6]. NeuAc has also been detected as a surface component of a number of pathogenic microorganisms [4,5,7-9], where it is thought to function in molecular mimicry of host glycoforms and in immune evasion.

Previously, our lab [2,3] reported the characterization of chimeric lipopolysaccharides (LPS) from *E. coli* K-12 transformed with lipooligosaccharide synthesis genes (*lsg*) from capsular *Haemophilus influenzae* type b (Hib). The expression of exogenous LPS was dependent on activation of the *E. coli wecA* gene by the LsgG gene product. Chimeric oligosaccharides synthesized with this system were linked to the 7-position of the non-reducing terminal branch heptose of the complete *E. coli* LPS core structure. We have used this system to investigate the expression of the entire *lsg* gene locus (vector pGEM-LOS-04), as well as a truncated version of the *lsg* gene locus (vector pGEM-LOS-05) [3]. Table 4-2 features the *lsg* genes present in these studies and the resulting carbohydrate structures synthesized onto the *E. coli* LPS core.

***Table 4- 1: Chimeric oligosaccharides from previous chimeric studies***

| Phillips et al. Transformant | Relevant Hib ORF | Chimeric Oligosaccharide |
|---|---|---|
| PGEM-LOS-05 | *lsgC, D, E, F, G* | Gal1–4 GlcNAc1–3 Gal1–3 GlcNAc→CORE |
| PGEM-LOS-04 | *lsgA, B, C, D, E, F, G* | GlcNAc/Gal1–6 Gal1–4 GlcNAc1–3 Gal1–3 GlcNAc→CORE |

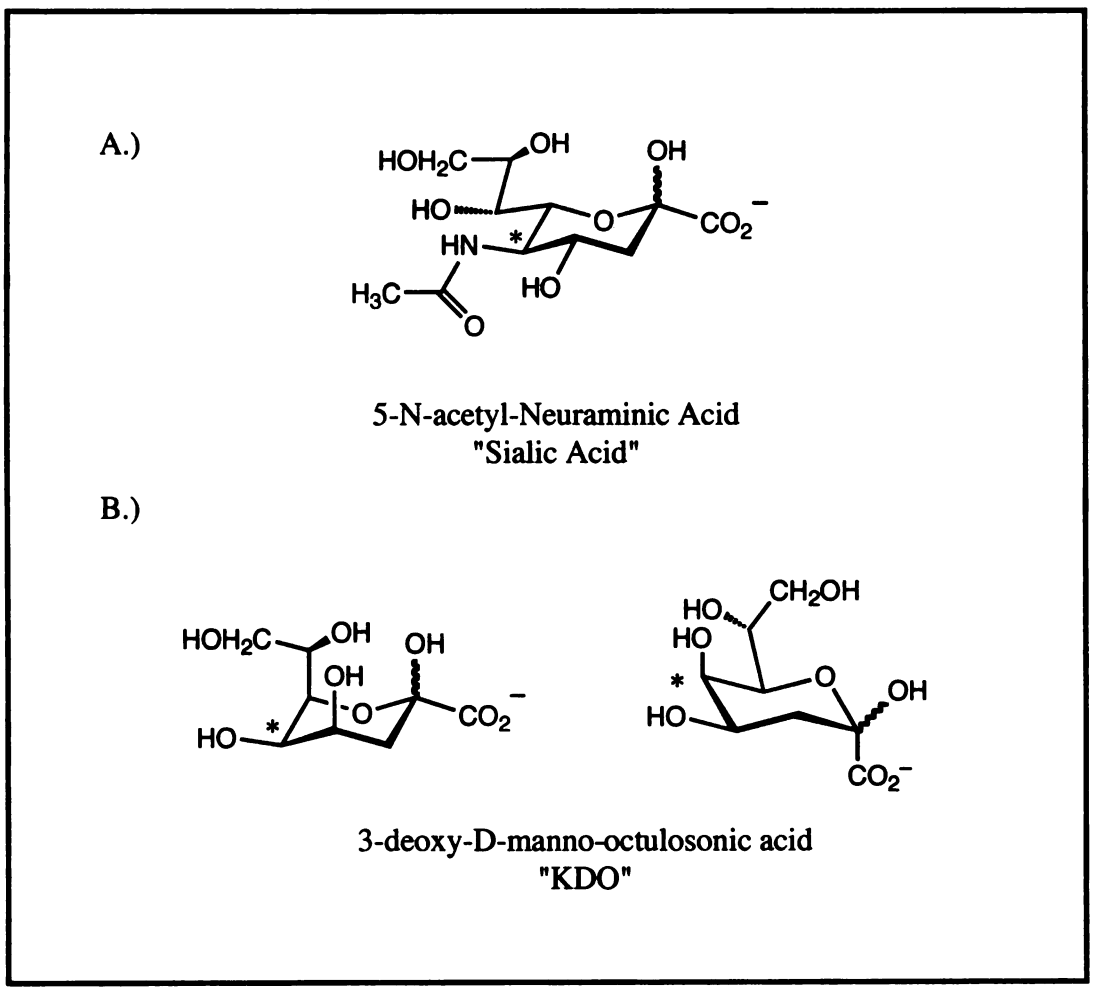
In the previous chapter, I described the chimeric expression system to express the four pGEM-LOS-05 glycosyltransferases *lsgC, D, E,* and *F*, individually, and in two- and three-gene combinations. It was determined that the gene products LsgD, E, and F synthesize the chimeric "LOS-05" tetrasaccharide [2] in the presence of LsgG. The functions of the *lsg* locus gene products LsgA, B, and C have yet to be determined. LsgA has homology to an O-antigen transporter or flippase, LsgB has high homology to an α-2,3 sialyltransferase, and LsgC has homology to a glycosyltransferase [10].

The chimeric LPS from the LOS-04 transformant is one Galactose (Gal) or N-acetyl glucosamine (GlcNAc) larger than the chimeric LPS from the LOS-05 transformant. Additionally, both of these larger chimeric LPS structures also contain an extra 3-deoxy-D-manno-octulosonic acid (KDO). However, the only difference between the pGEM-LOS-04 vector and the pGEM-LOS-05 vector was the presence of a (putative) flippase and sialyltransferase, respectively. This does not directly account for the presence of additional GlcNAc, Gal, and KDO residues.

One possibility was that the putative sialyltransferase LsgB required CMP-NeuAc, the presumed nucleotide-sugar substrate for α-2,3 sialyltransferases. The larger structures of the "LOS-04" chimeric LPS would then be the result of the presence of the putative LPS transporter LsgA. The larger oligosaccharide structures would be present in the pGEM-LOS-05 transformant, however the *E. coli* would be unable to export them.

An alternate possibility was that the LsgB gene product was capable of adding KDO to the terminus of the "LOS-05" chimeric LPS in the absence of CMP-NeuAc. In this case the addition of the KDO would be necessary to generate the nascent chain

***Figure 4-1: Structure of NeuAc and both conformational representations of KDO***




A.) Chemical structure of 5-N-acetyl-neuraminic acid (Neu5Ac, NeuAc).
B.) Chemical structures of both anomeric forms of 3-deoxy-D-manno-octulosonic acid (KDO). * Carbon 5 of NeuAc and KDO.

substrate for a Gal or GlcNAc transferase present in the pGEM-LOS-05 transformant. The structures of NeuAc and KDO are featured in Figure 4-1. A sialyltransferase that was not highly selective for substituents on carbon 5 of its nucleotide-sugar substrate could conceivably accept CMP-KDO as a substrate. Sialyltransferases that are tolerant of substitution at carbon 5 have been reported previously [11,12].

Efforts to establish the function of the LsgB gene product in its endogenous system have met with resistance because of degenerate sialylation pathways in *H. influenzae* [4] and the heterogeneity and variability of Hib LOS [7,13-15]. Expression of the *lsgB* sialyltransferase in an *E. coli* background containing CMP-NeuAc would provide a simpler system in which to evaluate its function.

In the current study, I adapt this chimeric expression system to express the CMP-NeuAc synthetase SiaB from NTHI. This should provide an opportunity to explore the synthetic properties (specificity and selectivity) of sialyltransferases such as LsgB, as well as develop an *in vivo* methodology for the synthesis of sialylated oligosaccharides and glycolipids.

In an effort to characterize the synthetic properties of the LsgA, B, and C, gene products for carbohydrate synthesis and better understand their role in Hib LOS biosynthesis, I report here the partial characterization of the LPS from *E. coli* strain XL-1 transformed with pGEM-LOS-05, pGEM-LOS-04 and pGEM-LOS-04 +pJCSiaB vectors via MALDI-TOF mass spectrometry.

## MATERIALS AND METHODS

***Bacterial Strains And Vectors:*** The *E. coli* K-12 Strain XL-1 Bluescript was obtained from Stratagene. Strains were routinely cultured at 37°C using LB agar or broth supplemented with the appropriate antibiotics. All vectors used in this study are listed in Table 2. The pGEM-LOS-04 and pGEM-LOS-05 vectors used in this study were previously described [3,16-18]. The vector pJCSiaB was obtained from our collaborators, the Apicella Lab at the University of Iowa [1].

***Table 4-2: Bacterial Strains and Vectors***

| Strain/Plasmid | Relevant Characteristics | Ref./Source |
|---|---|---|
| E. Coli | | |
| XL-1 Bluescript | *recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F'*proAB lacI*$^q$*ZDM15 Tn10* (Tet$^r$)] | Stratagene |
| Plasmids | | |
| pGEM-LOS-04 | Ap$^r$, contains 7.4-kb BamH I-Pst I DNA from *H. influenzae lsg locus* | [3] |
| pGEM-LOS-05 | Ap$^r$, contains 5.5-kb Hind III-Pst I DNA from *H. influenzae lsg locus* | [3] |
| pJCSiaB | Cm$^r$, contains ORF *siaB from N. meningitidis* | [1] |

***Cloning: Construction of Two-Vector Strain:*** To create the two-vector transformant pGEM-LOS-04 + pJCSiaB, *E. coli* containing the pJCSiaB vector was cultured overnight in 20 ml of LB supplemented with chloramphenicol. The following day, a 5 ml aliquot of the overnight culture was added to 60 ml LB:chloramphenicol and the bacteria was grown with shaking at 37°C until the bacteria reached an OD600 of 0.6. The bacteria were harvested by centrifugation at 5400 x *g* in a Jouan MR1812 centrifuge. The pellet

was washed three times with 50 ml of ice-cold, sterile, deionized water and resolubilized in 160 $\mu$l of ice-cold sterile deionized water. 40 $\mu$l of the now-electrocompetent pJCSiaB vector containing bacteria were transformed with the pGEM-LOS-04 vector using a Bio-Rad MicroPulser electroporator. The electroporation cocktail was plated onto LB agar plates supplemented with ampicillin and chloramphenicol. Qiagen mini-plasmid preps of the successful transformants were then screened via PCR for the presence of their respective vectors. pJCSiaB has an origin of replication that is compatible with the colE1 origin of the pGEM-LOS-04 vector.

***Table 4-3: Chimeric E. coli Transformants***

| Transformant | Vector 01 | Vector 02 | Relevant Hib ORFs |
|---|---|---|---|
| LOS-05 | pGEM-LOS-5 | | *lsgC, lsgD, lsgE, lsgF, lsgG* |
| LOS-04 | pGEM-LOS-4 | | *lsgA, lsgB, lsgC, lsgD, lsgE, lsgF, lsgG* |
| LOS-04 + SiaB | pGEM-LOS-4 | PJCSiaB | *lsgA, lsgB, lsgC, lsgD, lsgE, lsgF, lsgG, siaB* |

***LPS Extraction:*** The LPS of all strains were prepared using the modified hot phenol extraction procedure [19,20]. An additional DNAse incubation was performed to reduce recombinant DNA contamination from the vectors used to confer the LPS phenotypes. Briefly, bacteria was harvested by centrifugation and suspended in 2.0 ml of phosphate-buffered saline (PBS). The suspension was then centrifuged at 10,000 x *g* for 15 min at 4°C and the supernatant discarded. The pellet was washed once more with PBS and the resulting pellet was suspended in 600 $\mu$l of deionized water and incubated in a 65°C water bath for 15 min. An equal volume of hot (65°C) 90% phenol was added and the

mixture was incubated at 65°C for 35 min with frequent vortexing. The suspension was chilled on ice and then centrifuged at 8500 x *g* for 20 min, 4°C. The supernatant was transferred to a 15 ml conical tube, and the phenol fraction was re-extracted with 600 μl of deionized water. The aqueous phases were pooled, and 1/10 sample volume of 3 M sodium acetate pH 5.2 and three volumes of ethanol were added. The samples were then placed at –20°C overnight to precipitate the LPS. The precipitate was centrifuged at 10,000 x *g*, 4°C for 15 min and the supernatant discarded. The sample was resolubilized in 200 μl deionized water and re-precipitated. The precipitate was dried and resolubilized in 500 μl of 60 mM Tris-Cl (pH 8.0) + 2 mM $MgCl_2$, and 2 μg/ml of DNAse (Roche Applied Research) were added and the samples were incubated at 37°C for 2 h. The samples were then precipitated overnight as described above and the resulting pellets were re-solubilized in 1.5 ml of deionized water. The samples were then ultracentrifuged in a TL-100 ultracentrifuge (Beckman) at 120,000 x *g* for 2 h. The supernatants were discarded and the pellets were re-solubilized in 200 μl of deionized water and evaporated to dryness in a Speed-Vac concentrator (Savant). The dried samples were stored at –20°C.

***O-Deacylation of LPS:*** To make LPS samples more amenable to mass spectrometric analysis, LPS samples were treated with anhydrous hydrazine to remove their O-linked lipid chains. Briefly: dried samples of LPS (~100 μg) were treated with 100 μl of anhydrous hydrazine (Sigma) and incubated at 37°C for 30 min in microcentrifuge tubes with occasional vortexing [21]. After the reaction, the samples were placed in an ice bath and 500 μl of chilled acetone was added dropwise to precipitate the O-deacylated

LPS (O-LPS). Samples were allowed to sit at –20°C for about 1 h and then they were centrifuged at 12,000 x *g* at 4°C. The supernatants were removed and the O-deacylated LPS pellets were dissolved in 100 μl of deionized water and evaporated to dryness on a Speed-Vac concentrator.

***Sample Preparation for MALDI:*** To remove salts and other low-molecular-weight contaminants from O-LPS the samples were drop dialyzed. Approximately 100 μg of dry sample was dissolved in 20 μl of deionized water. Then 10 μl aliquots were desalted by drop dialysis on VSWP 0.025 μm nitrocellulose membranes (Millipore) over deionized water for 2 h. Samples were recovered from the membranes, evaporated to dryness, and redissolved in 10 μl of deionized water for mass spectrometric analysis.

***MALDI-TOF Mass Spectrometry:*** To determine molecular weights of O-LPS samples, they were analyzed on a Voyager MALDI-TOF instrument (Applied Biosystems) equipped with a nitrogen laser (337 nm). All spectra were recorded in the negative-ion mode using delayed extraction conditions with an accelerating voltage of 20,000 V. Briefly: a one μl portion containing ~10 μg of a dialyzed O-LPS sample was delivered into a 0.5 ml microcentrifuge tube containing a few beads of Dowex-50W-X8 resin (100-200 mesh, $NH_4^+$ form) and allowed to sit for 1 min. One μl of the matrix solution (a saturated solution of 2,5-dihydroxybenzoic acid in acetone) was then added to the sample and mixed briefly. Finally, the samples were spotted on a stainless steel MALDI target and allowed to air-dry. Approximately 150 laser shots were recorded and averaged for each sample. The spectra were smoothed with a 19-point Savitsky-Golay function and

mass-calibrated using an external mass calibrant consisting of renin substrate tridecapeptide, oxidized insulin chain B, and bovine insulin (Sigma).

## RESULTS

Negative ion MALDI-TOF spectra of the O-deacylated LPS (O-LPS) from *E. coli* transformants LOS-05, LOS-04, and LOS-04 + SiaB (vectors and expressed ORFs of each "transformant" are detailed in Table 4-3) are shown in Figure 4-2. The *m/z* values of each peak were matched to calculated *m/z* values from proposed compositions, and this data is featured in Table 4-4.

Two things are apparent from looking at the stack plot of spectra in Figure 4-2. First is that there are molecular ions present in the MALDI spectrum of the O-LPS of the LOS-04 + SiaB transformant that are much larger than any of the molecular ions of the other two transformants. Secondly, it is apparent that the molecular ions in the mass range between *m/z* 2500-3500, which correspond to the *E. coli* LPS core glycoforms, are more highly abundant relative to the molecular ions in the mass range between *m/z* 3500-4500, which correspond to the chimeric LPS extension glycoforms, in the MALDI spectrum of the LOS-04 + SiaB O-LPS. In the spectra for both of the other transformants the relative abundance of molecular ions between these two pools of glycoforms is relatively similar.

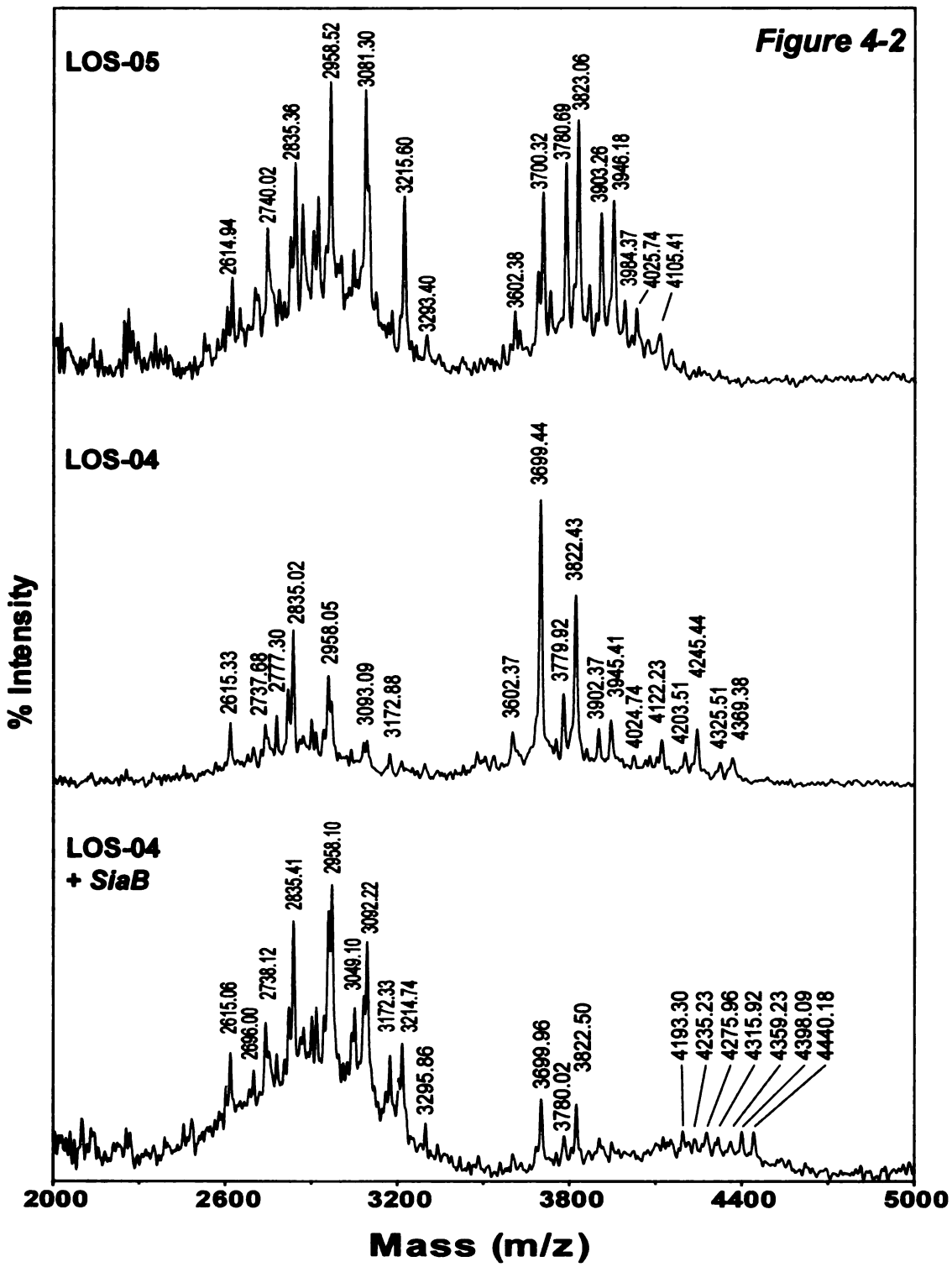



***Figure 4-2***: Negative ion MALDI-TOF spectra of the O-deacylated LPS from *E. coli* K-12 transformants LOS-05, LOS-04, and LOS-04+*SiaB*. Each glycoform exists as a mixture of phosphoforms and phosphoethanolamine-forms. Peaks are annotated in Table 4-3.

***Table 4-4: Molecular weights and proposed compositions of the O-deacylated LPS from E. coli K-12 transformants* LOS-05, LOS-04, and LOS-04 + SiaB***

| Transformant: | LOS-05 | LOS-04 | LOS-04+SiaB | Calculated | Proposed Compositions‡ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Glycoform | Experimental Average *m/z* (M-H)⁻ | | | *m/z* (M-H)⁻ | Hex | Hep | HexNac | PO4 | PEA | KDO | NeuAc | O-DPLA |
| E5B+NeuAc | | | 4440.2 | 4440.8 | 6 | 4 | 3 | 2 | 2 | 2 | 1 | 1 |
| E5B+NeuAc | | | 4315.9 | 4317.7 | 6 | 4 | 3 | 2 | 1 | 2 | 1 | 1 |
| E5B+NeuAc | | | 4193.3 | 4194.7 | 6 | 4 | 3 | 2 | | 2 | 1 | 1 |
| E5A+NeuAc | | | 4398.1 | 4399.7 | 7 | 4 | 2 | 2 | 2 | 2 | 1 | 1 |
| E5A+NeuAc | | | 4359.2 | 4356.7 | 7 | 4 | 2 | 3 | 1 | 2 | 1 | 1 |
| E5A+NeuAc | | | 4276.0 | 4276.7 | 7 | 4 | 2 | 2 | 1 | 2 | 1 | 1 |
| E4+NeuAc | | | 4235.2 | 4237.6 | 6 | 4 | 2 | 2 | 2 | 2 | 1 | 1 |
| E5B+KDO | | 4369.4 | | 4369.7 | 6 | 4 | 3 | 2 | 2 | 3 | | 1 |
| E5B+KDO | | 4245.4 | | 4246.7 | 6 | 4 | 3 | 2 | 1 | 3 | | 1 |
| E5B+KDO | | 4122.2 | | 4123.6 | 6 | 4 | 3 | 2 | | 3 | | 1 |
| E5A+KDO | | 4325.5 | | 4328.7 | 7 | 4 | 2 | 2 | 2 | 3 | | 1 |
| E5A+KDO | | 4203.5 | | 4205.6 | 7 | 4 | 2 | 2 | 1 | 3 | | 1 |
| E4 | 4105.4 | | | 4106.3 | 6 | 4 | 2 | 4 | 2 | 2 | | 1 |
| E4 | 4025.7 | 4024.7 | | 4026.3 | 6 | 4 | 2 | 3 | 2 | 2 | | 1 |
| E4 | 3984.4 | | | 3983.2 | 6 | 4 | 2 | 4 | 1 | 2 | | 1 |
| E4 | | 3945.4 | | 3946.3 | 6 | 4 | 2 | 2 | 2 | 2 | | 1 |
| E4 | 3903.3 | 3902.4 | | 3903.3 | 6 | 4 | 2 | 3 | 1 | 2 | | 1 |
| E4 | 3823.1 | 3822.4 | 3822.5 | 3823.3 | 6 | 4 | 2 | 2 | 1 | 2 | | 1 |
| E4 | 3780.7 | 3779.9 | 3780.0 | 3780.2 | 6 | 4 | 2 | 3 | | 2 | | 1 |
| E4 | 3700.3 | 3669.7 | 3700.0 | 3700.2 | 6 | 4 | 2 | 2 | | 2 | | 1 |
| E4 | 3602.4 | 3602.4 | 3602.3 | 3603.1 | 6 | 4 | 2 | 2 | 1 | 1 | | 1 |
| C2 | 3293.4 | | 3295.9 | 3295.6 | 4 | 4 | | 3 | 2 | 2 | | 1 |
| C2 | 3215.6 | | 3214.7 | 3215.7 | 4 | 4 | | 2 | 2 | 2 | | 1 |
| C2 | 3172.7 | 3172.9 | 3172.3 | 3172.6 | 4 | 4 | | 3 | 1 | 2 | | 1 |
| C2 | 3091.4 | 3093.1 | 3092.2 | 3092.6 | 4 | 4 | | 2 | 1 | 2 | | 1 |
| C2 | | | 3049.1 | 3049.5 | 4 | 4 | | 3 | | 2 | | 1 |
| C1 | 3081.3 | 3081.7 | 3081.0 | 3081.5 | 3 | 3 | | 2 | 2 | 3 | | 1 |
| C1 | 2958.5 | 2958.1 | 2958.1 | 2958.5 | 3 | 3 | | 2 | 1 | 3 | | 1 |
| C1 | 2942.8 | | | 2941.3 | 3 | 3 | | 3 | 2 | 2 | | 1 |
| C1 | 2915.6 | | | 2915.4 | 3 | 3 | | 3 | | 3 | | 1 |
| C1 | 2861.5 | | | 2861.3 | 3 | 3 | | 2 | 2 | 2 | | 1 |
| C1 | 2835.4 | 2835.0 | 2835.4 | 2835.4 | 3 | 3 | | 2 | | 3 | | 1 |
| C1 | 2740.0 | 2737.7 | 2738.1 | 2738.3 | 3 | 3 | | 2 | 1 | 2 | | 1 |
| C1 | 2778.8 | 2777.3 | 2776.9 | 2775.2 | 3 | 3 | | 4 | | 2 | | 1 |
| C1 | 2694.8 | 2696.6 | 2696.0 | 2695.2 | 3 | 3 | | 3 | | 2 | | 1 |
| C1 | 2614.9 | 2615.3 | 2615.1 | 2615.3 | 3 | 3 | | 2 | | 2 | | 1 |

*The vector contents of each *E. coli* K-12 transformant are detailed in Table 4-3. ‡The residue mass values (average mass) for LPS structural moities are as follows: Hex (162.142), Hep (192.169), HexNAc (203.195), phosphate (79.980), PEA (123.048), KDO (220.179), NeuAc (291.258), and diphosphorylated O-deacylated Lipid A (934.994).

In all three spectra, the molecular ions in the mass range between *m/z* 2500-3500 correspond to the *E. coli* core LPS glycoforms, and are labeled as the C1 and C2 glycoforms in Table 4-4. All three spectra also feature molecular ions between *m/z* of 3500 and 4150. These ions correspond to tetrasaccharide extensions of the *E. coli* core LPS glycoforms, and are labeled as the E4 glycoforms in Table 4-4.

In the spectrum of the O-LPS from the LOS-04 transformant, there is a grouping of molecular ions between *m/z* 4120-4400. These ions correspond to pentasaccharide extensions of the *E. coli* core LPS glycoforms, which also feature an additional KDO. These ions are labeled as the E5A + KDO and E5B + KDO glycoforms in Table 4-4.

In the spectrum of the O-LPS from the "LOS-04 + SiaB" transformant, there is a set of molecular ions between *m/z* 4190-4450. These ions correspond to pentasaccharide extensions of the *E. coli* core LPS glycoforms, which also feature an additional NeuAc. These ions are labeled as the E5A + NeuAc and E5B + NeuAc glycoforms in Table 4-3.

In all three spectra, each of the observed glycoforms features a moderate degree of phosphate ($PO_4$) and phosphoethanolamine (PEA) heterogeneity. However, the only glycoform that seems to exhibit KDO heterogeneity is the C1 glycoform. Neither the C2 or E4 glycoforms have any observable KDO heterogeneity. All of the E5A glycoforms present in the O-LPS of the LOS-04 transformant that feature an additional KDO have the same number of KDOs. It is possible that the addition of a third KDO to the C-1 glycoform prevents the further extension of this glycoform, although the mechanism for this addition remains unknown.

It is interesting to note that the relative abundance of each phospho-form and phosphoethanolamine-form of the E5A and E5B LPS glycoforms is very similar between

the LPS of the LOS-04 transformant and the LOS-04 + SiaB transformant. The only exception being that the E5A and E5B glycoforms contain an additional NeuAc in the LOS-04 + SiaB transformant instead of an extra KDO. No E5A or E5B glycoform ions could be detected in the spectrum of the O-LPS from the LOS-04 or LOS-04 + SiaB transformants that contain both an additional KDO and an additional NeuAc.

## DISCUSSION

The fact that we have observed sialylated glycoforms in the LOS-04 + SiaB transformant confirms the hypothesis that the chimeric LPS expression system can be adapted to the synthesis of sialylated glycoforms by the addition of a CMP-NeuAc synthetase. Furthermore we have also confirmed that LsgB is indeed a sialyltransferase. The fact that all of the sialylated LPS glycoforms observed in the LOS-04 + SiaB transformant were lacking the additional KDO observed in the LPS of the LOS-04 lends support to the hypothesis that LsgB is capable of recognizing KDO as a substrate in the absence of CMP-NeuAc. It is possible that the additional Gal and GlcNAc additions of the E5 glycoforms can only occur once a KDO or NeuAc has been added to the E4 glycoform.

## CONCLUSIONS

The *E. coli* chimeric LPS synthesis system can be adapted for the synthesis of sialylated glycoforms. A biosynthetic or bacterial metabolism engineering approach to the synthesis of oligosaccharides and glycolipids is especially advantageous in the case of synthesizing sialylated glycans. Unlike synthetic and even *in vitro* enzymatic synthetic methodologies, there is no need to procure CMP-NeuAc which can be unstable and costly [22,23]. We have developed a system in which sialylated oligosaccharides and glycolipids can be synthesized in the same *E.coli* bacteria, without the need for engineering synthetase-transferase fusion proteins or complicated systems of bacterial coupling.

Furthermore, I have developed a simplified system for profiling the synthetic function of putative microbial sialyltransferases. Sialylation pathways are extremely complicated and feature extensive degeneracy [4,18,22-24]. Also, many of the host species are known to phase vary their surface carbohydrate epitopes and sialylation patterns. The ability to profile the function of putative microbial sialyltransferases may yet prove to be this system's greatest potential.

## REFERENCES


1. Greiner LL, Watanabe H, Phillips NJ, Shao J, Morgan A, Zaleski A, Gibson BW, Apicella MA: **Nontypeable Haemophilus influenzae strain 2019 produces a biofilm containing N-acetylneuraminic acid that may mimic sialylated O-linked glycans.** *Infect Immun* 2004, **72**:4249-4260.
2. Johansen E, Gibson BW: **Functional Exploration of the *lsg* gene locus glycosyltransferases *lsgC, D, E, and F,* through chimeric expression in Eschericia coli K-12.** *Manuscript in Preparation* 2006.
3. Phillips NJ, Miller TJ, Engstrom JJ, Melaugh W, McLaughlin R, Apicella MA, Gibson BW: **Characterization of chimeric lipopolysaccharides from Escherichia coli strain JM109 transformed with lipooligosaccharide synthesis genes (lsg) from Haemophilus influenzae.** *J Biol Chem* 2000, **275**:4747-4758.
4. Jones PA, Samuels NM, Phillips NJ, Munson RS, Jr., Bozue JA, Arseneau JA, Nichols WA, Zaleski A, Gibson BW, Apicella MA: **Haemophilus influenzae type b strain A2 has multiple sialyltransferases involved in lipooligosaccharide sialylation.** *J Biol Chem* 2002, **277**:14598-14611.
5. Vimr ER, Kalivoda KA, Deszo EL, Steenbergen SM: **Diversity of microbial sialic acid metabolism.** *Microbiol Mol Biol Rev* 2004, **68**:132-153.
6. Schauer R: **Achievements and challenges of sialic acid research.** *Glycoconj J* 2000, **17**:485-499.
7. Swords WE, Jones PA, Apicella MA: **The lipo-oligosaccharides of Haemophilus influenzae: an interesting array of characters.** *Journal of Endotoxin Research* 2003, **9**:131-144.
8. Vimr E, Lichtensteiger C: **To sialylate, or not to sialylate: that is the question.** *Trends Microbiol* 2002, **10**:254-257.
9. Harvey HA, Swords WE, Apicella MA: **The mimicry of human glycolipids and glycosphingolipids by the lipooligosaccharides of pathogenic neisseria and haemophilus.** *J Autoimmun* 2001, **16**:257-262.
10. Schaffer AA, Aravind L, Madden TL, Shavirin S, Spouge JL, Wolf YI, Koonin EV, Altschul SF: **Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements.** *Nucleic Acids Res* 2001, **29**:2994-3005.
11. Takashima S, Abe T, Yoshida S, Kawahigashi H, Saito T, Tsuji S, Tsujimoto M: **Analysis of Sialyltransferase-Like Proteins from Oryza sativa.** *J Biochem (Tokyo)* 2006, **139**:279-287.
12. Schilling B, Goon S, Samuels NM, Gaucher SP, Leary JA, Bertozzi CR, Gibson BW: **Biosynthesis of sialylated lipooligosaccharides in Haemophilus ducreyi is dependent on exogenous sialic acid and not mannosamine. Incorporation studies using N-acylmannosamine analogues, N-glycolylneuraminic acid, and 13C-labeled N-acetylneuraminic acid.** *Biochemistry* 2001, **40**:12666-12677.

13. Phillips NJ, John CM, Reinders LG, Gibson BW, Apicella MA, Griffiss JM: **Structural models for the cell surface lipooligosaccharides of Neisseria gonorrhoeae and Haemophilus influenzae.** *Biomed Environ Mass Spectrom* 1990, **19**:731-745.
14. Phillips NJ, Apicella MA, Griffiss JM, Gibson BW: **Structural studies of the lipooligosaccharides from Haemophilus influenzae type b strain A2.** *Biochemistry* 1993, **32**:2003-2012.
15. Gaucher SP, Cancilla MT, Phillips NJ, Gibson BW, Leary JA: **Mass spectral characterization of lipooligosaccharides from Haemophilus influenzae 2019.** *Biochemistry* 2000, **39**:12406-12414.
16. Spinola SM, Kwaik YA, Lesse AJ, Campagnari AA, Apicella MA: **Cloning and expression in Escherichia coli of a Haemophilus influenzae type b lipooligosaccharide synthesis gene(s) that encodes a 2-keto-3-deoxyoctulosonic acid epitope.** *Infect Immun* 1990, **58**:1558-1564.
17. Abu Kwaik Y, McLaughlin RE, Apicella MA, Spinola SM: **Analysis of Haemophilus influenzae type b lipooligosaccharide-synthesis genes that assemble or expose a 2-keto-3-deoxyoctulosonic acid epitope.** *Mol Microbiol* 1991, **5**:2475-2480.
18. Mandrell RE, McLaughlin R, Aba Kwaik Y, Lesse A, Yamasaki R, Gibson B, Spinola SM, Apicella MA: **Lipooligosaccharides (LOS) of some Haemophilus species mimic human glycosphingolipids, and some LOS are sialylated.** *Infect Immun* 1992, **60**:1322-1328.
19. Inzana TJ: **Electrophoretic heterogeneity and interstrain variation of the lipopolysaccharide of Haemophilus influenzae.** *J Infect Dis* 1983, **148**:492-499.
20. Apicella MA, Griffiss JM, Schneider H: **Isolation and characterization of lipopolysaccharides, lipooligosaccharides, and lipid A.** *Methods Enzymol* 1994, **235**:242-252.
21. Helander IM, Lindner B, Brade H, Altmann K, Lindberg AA, Rietschel ET, Zahringer U: **Chemical structure of the lipopolysaccharide of Haemophilus influenzae strain I-69 Rd-/b+. Description of a novel deep-rough chemotype.** *Eur J Biochem* 1988, **177**:483-492.
22. Endo T, Koizumi S, Tabata K, Ozaki A: **Large-scale production of CMP-NeuAc and sialylated oligosaccharides through bacterial coupling.** *Appl Microbiol Biotechnol* 2000, **53**:257-261.
23. Gilbert M, Bayer R, Cunningham AM, DeFrees S, Gao Y, Watson DC, Young NM, Wakarchuk WW: **The synthesis of sialylated oligosaccharides using a CMP-Neu5Ac synthetase/sialyltransferase fusion.** *Nat Biotechnol* 1998, **16**:769-772.
24. Antoine T, Priem B, Heyraud A, Greffe L, Gilbert M, Wakarchuk WW, Lam JS, Samain E: **Large-scale in vivo synthesis of the carbohydrate moieties of gangliosides GM1 and GM2 by metabolically engineered Escherichia coli.** *Chembiochem* 2003, **4**:406-412.

# CHAPTER 5.

# CHIMERIC CONCLUSIONS AND FUTURE EXPERIMENTS

## CONCLUSIONS

The work contained within this dissertation lays the groundwork for a modular system in which to express microbial glycosyltransferases for the purposes of determining their synthetic properties. At the outset of this project, four major hurdles existed between the intent and the actualization of this proposal. First, a modular system with ability to shuffle glycosyltransferase genes in a way that could lead to the synthesis of defined one and two-sugar nascent oligosaccharide chains had not been developed. Second, the methods of glycolipid extraction and purification being used when this project began were designed to extract lipooligosaccharides (LOS) from organisms like *Hamophilus ducreyi, Haemophilus influenzae*, and *Neisseria gonorrhoeae*. These were strains that did not typically contain large amounts of plasmid DNA. Also, the LOS extracted from these bacteria were typically much smaller and less heterogeneous than the lipopolysaccharides (LPS) routinely extracted from the chimeric *E. coli* bacteria in this work. The LPS extraction and purification protocols developed in the process of this work (see Appendix) will prove to be extremely useful to the future of this project. Third, I have determined the genes required for synthesis of a small number of defined one two- and four-sugar nascent oligosaccharide chains, with which to begin the next round of glycosyltransferase functional exploration. Figure 5-1A and 5-1B depict the chemical structures of each of the carbohydrate backgrounds synthesized thus far.

*Figure 5-1A: Chimeric LPS structures synthesized as part of this work*

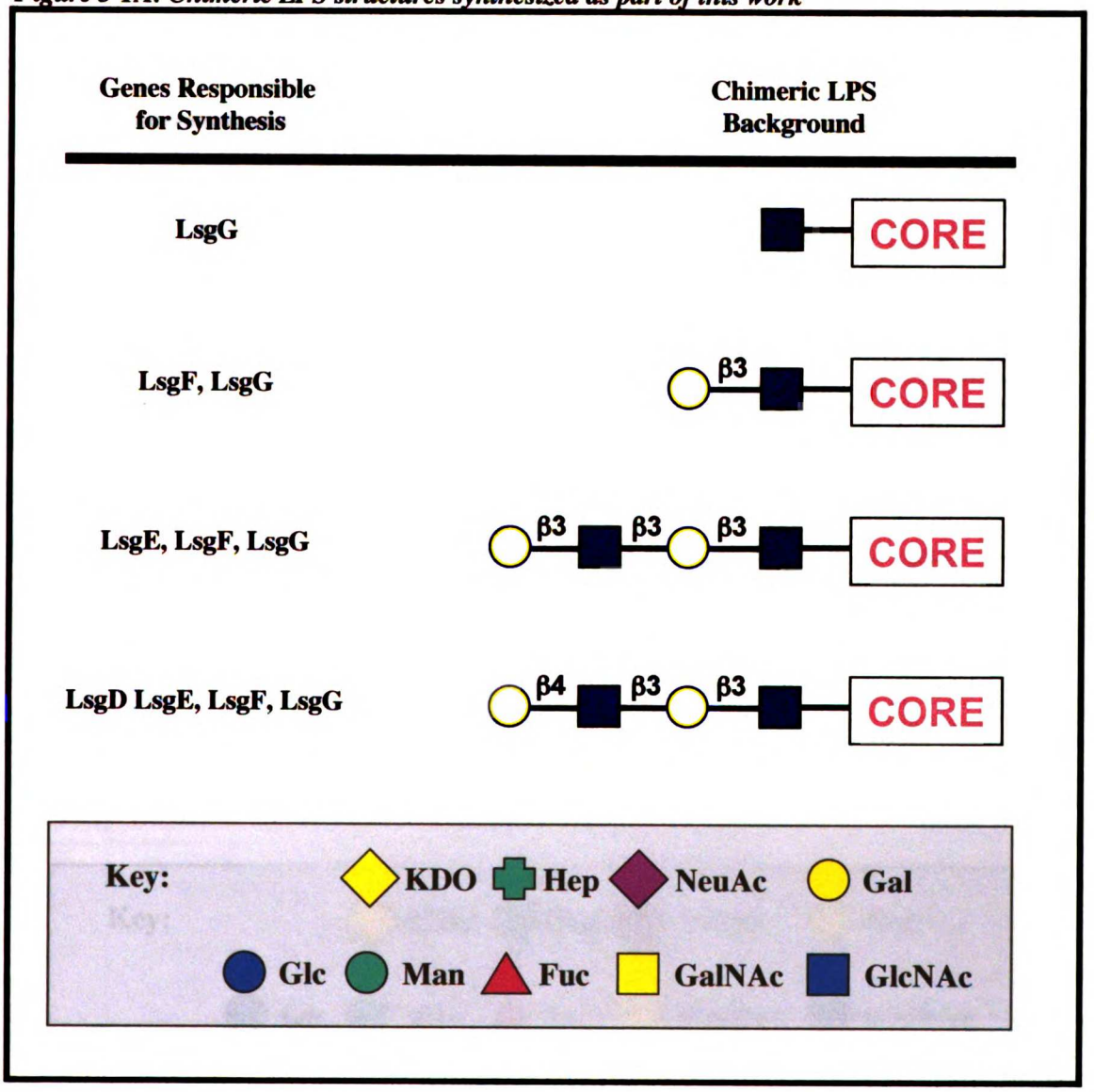
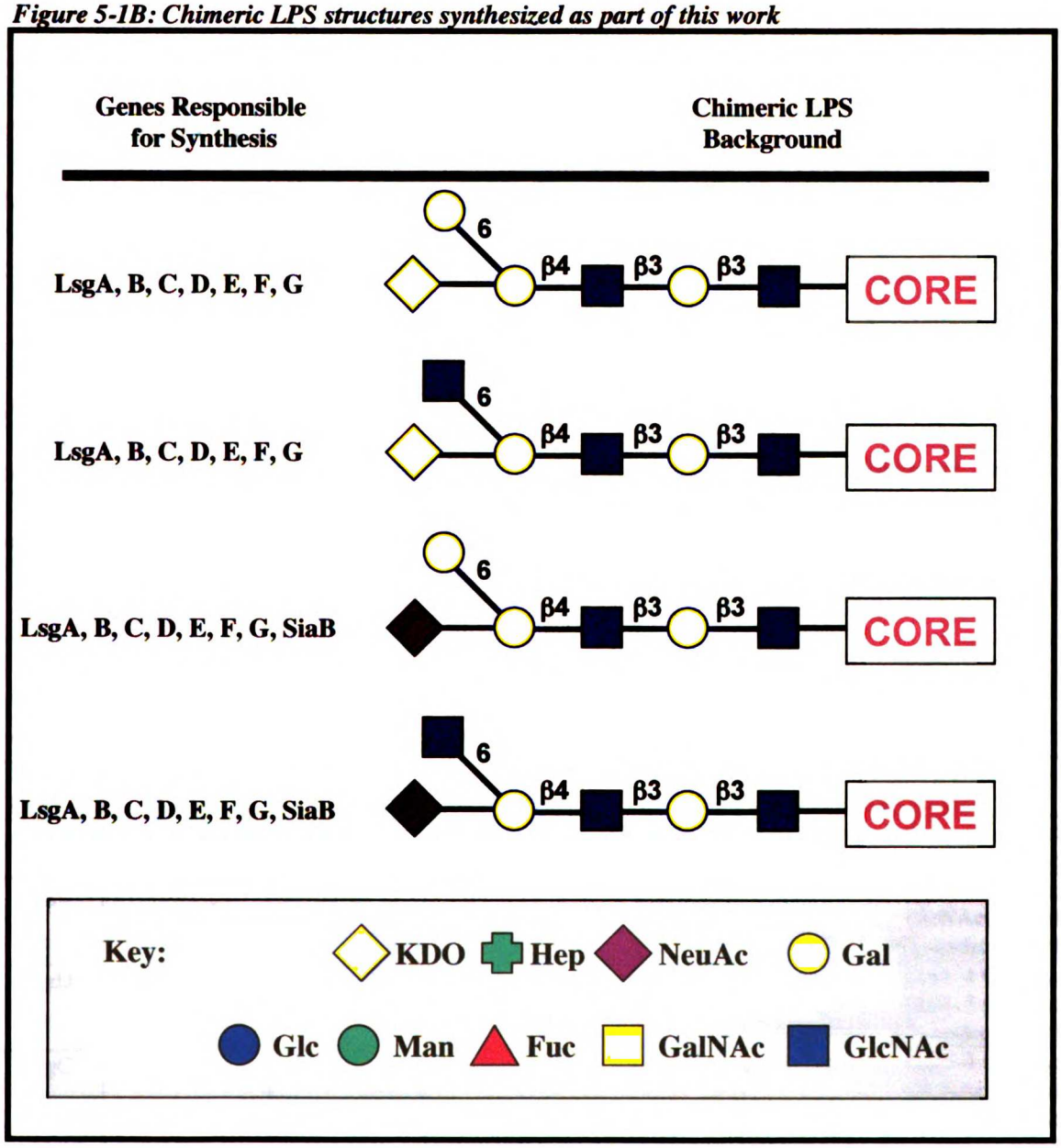


*Figure 5-1B: Chimeric LPS structures synthesized as part of this work*



Fourth, I had not yet ascertained whether one could adapt this chimeric system for the functional determination of sialyltransferases. Now, the function of any sialyltransferase capable of using CMP-Neu5Ac as a substrate can theoretically be determined using this system.

The data determined in this work regarding the synthetic properties of five glycosyltransferases: LsgB, LsgC, LsgD, LsgE and LsgF, can be readily implemented for a chemoenzymatic or bacterial metabolic engineering method of complex carbohydrate synthesis [1]. The function of these glycosyltransferases has previously eluded assignment [2-8]. A brief summary of our conclusions regarding the synthetic properties of these enzymes is shown in Table 5-1. Knowledge of the synthetic properties of these *lsg* gene products will also assist in the elucidation of the molecular mechanisms of Hib LOS synthesis and its contribution to pathogenicity.

***Table 5-1: Specificity and Selectivity Conclusions***

| Gene Product | Monomer Added | Linkage added | Nascent Chain Disaccharides Extended | Nascent Chain Disaccharides not Extended |
|---|---|---|---|---|
| LsgF | Galactose | 1–3 | GlcNAc1–3 Gal<br>GlcNAc1– $PO_4$-undecaprenol | Gal1–4 GlcNAc<br>Gal1–3 GlcNAc |
| LsgE | N-Acetyl Glucosamine | 1–3 | Gal1–3GlcNAc | Gal1–4 GlcNAc<br>GlcNAc1–3 Gal<br>GlcNac1– $PO_4$-undecaprenol |
| LsgD | Galactose | 1–4 | GlcNAc1–3Gal | Gal1–4 GlcNAc<br>Gal1–3 GlcNAc<br>GlcNac1– $PO_4$-undecaprenol |
| LsgC | ND | ND | | Gal1–4 GlcNAc<br>GlcNAc1–3 Gal<br>Gal1–3 GlcNAc<br>GlcNac1– $PO_4$-undecaprenol |
| LsgB | Neu5Ac<br>KDO(?) | ND | Gal1–4 GlcNAc | ND |

Additionally, these genes have been successfully cloned and have been used to express oligosaccharide. I have synthesized the lacto-N-neotetraose (Gal1–4-GlcNAc1–3Gal1–3GlcNAc) glycoform [9], using this chimeric expression system, a carbohydrate that is typically associated with *Neisseria gonorrhoeae* and *Neisseria meningitidis* LOS [9,10]. This tetrasaccharide has been implicated in the invasion of human urethral epithelial cells [9,11]. Large-scale preparation of LPS from transformant DEFG could potentially be a safer and more reliable way to obtain large quantities of this lipid-sugar. *N. gonorrhoeae* and *N. meningitidis* are both human pathogens [12] whereas transformant DEFG is not. Furthermore, *N. gonorrhoeae* and *N. meningitidis* LOS have been shown to demonstrate a high degree of phase variation [12], whereas transformant DEFG does not.

## FUTURE EXPERIMENTS

For this multiplasmid expression system to be expanded as a general method for the functional assignment of microbial glycosyltransferases, additional work will be required. I have developed a system with the ability to display synthesized glycans on its outer membrane. I have also developed the ability to construct combinations of glycosyltransferase gene-containing vectors, as well as the ability to express sialylated glycoforms. In the future, this system should be scaled up to include larger numbers of glycosyltransferase genes. For example, five unknown glycosyltransferases could be cloned into pACYC184 vectors featuring a chloramphenicol resistance cassette. The sets

of glycosyltransferase genes known to work together to produce a relevant backgrounds from this work (Figure 5-1A/B) could be cloned into a pGEM-LOS-12 vector featuring ampicillin resistance and the *lsgG* ORF. Screening the LPS from each of the resulting 25 colonies using traditional growth, extraction, purification, chemical modification and mass spectrometric analysis methods would be quite cumbersome. However, in an expression system where the carbohydrate products are expressed on the outer membrane it would be a simple matter to screen the whole cell bacteria with an antibody against a carbohydrate epitope of interest. All that would be required would be to grow control colonies containing only the pGEM-LOS-12 vector library in parallel with the pGEM-LOS-12 + pACYC184 vector library. Elisa antibody screening with a small number of available carbohydrate antibodies should permit detection of phenotypic variation between control and variable experimental groupings.

I believe the next step in the successful execution of this chimeric carbohydrate synthesis would be to profile a number of *E. coli* strains that are known to express truncated core LPS glycoforms with this expression system. We have determined that the only *E. coli* core LPS extended by our system is the one featuring terminal non-reducing L-glycero-D-mannoheptose [4,13]. However, there may yet exist systems that feature a similar terminal heptose, but are unable to add PEA residues, or are deficient in phosphorylation of the LPS. The core LPS of *E. coli* K-12 is simply too large and too heterogeneous for simple extraction, isolation, and profiling. All advances within this system have come at great cost of time and frustration. That does not mean that this system holds no potential, however. The benefits of this system far outweigh the

drawbacks, and would doubly so if a new chimeric expression host could be found with a much simpler and less heterogeneous core LPS molecule.

As this chimeric synthesis methodology continues to be optimized and expanded to include more glycosyltransferase genes it may well have its greatest potential as a general method for the production of specific oligosaccharides and lipid-linked oligosaccharides. One particularly exciting application of this technology might be the construction of carbohydrate-based vaccines, where carbohydrate epitopes from pathogenic bacteria could be incorporated into non-lethal strains. These strains could comprise the vaccine itself. Alternatively, there have been advances in the field of vaccine adjuvants based on the structure of Lipid A, such as monophosphoryl Lipid-A [14]. Biosynthetic chimeric LPS representing a carbohydrate vaccine of interest could possibly be dephosphorylated and used as a vaccine without the need for additional adjuvants.

## REFERENCES


1. Sears P, Wong CH: **Toward automated synthesis of oligosaccharides and glycoproteins.** *Science* 2001, **291**:2344-2350.
2. Hood DW, Randle G, Cox AD, Makepeace K, Li J, Schweda EKH, Richards JC, Moxon ER: **Biosynthesis of Cryptic Lipopolysaccharide Glycoforms in Haemophilus influenzae Involves a Mechanism Similar to That Required for O-Antigen Synthesis.** *J. Bacteriol.* 2004, **186**:7429-7439.
3. Swords WE, Jones PA, Apicella MA: **The lipo-oligosaccharides of Haemophilus influenzae: an interesting array of characters.** *Journal of Endotoxin Research* 2003, **9**:131-144.
4. Phillips NJ, Miller TJ, Engstrom JJ, Melaugh W, McLaughlin R, Apicella MA, Gibson BW: **Characterization of chimeric lipopolysaccharides from Escherichia coli**

**strain JM109 transformed with lipooligosaccharide synthesis genes (lsg) from Haemophilus influenzae.** *J Biol Chem* 2000, **275**:4747-4758.

5. Phillips NJ, McLaughlin R, Miller TJ, Apicella MA, Gibson BW: **Characterization of two transposon mutants from Haemophilus influenzae type b with altered lipooligosaccharide biosynthesis.** *Biochemistry* 1996, **35**:5937-5947.
6. McLaughlin RE, Lee N-G, Abu Kwaik Y, Spinola SM, Apicella MA: **Characterization and sequence analysis of the lsg (LOS synthesis genes) locus from *Haemophilus influenzae* type b.** *J. Endotoxin Res.* 1994, **1**:165-174.
7. Phillips NJ, Apicella MA, Griffiss JM, Gibson BW: **Structural studies of the lipooligosaccharides from Haemophilus influenzae type b strain A2.** *Biochemistry* 1993, **32**:2003-2012.
8. Spinola SM, Kwaik YA, Lesse AJ, Campagnari AA, Apicella MA: **Cloning and expression in Escherichia coli of a Haemophilus influenzae type b lipooligosaccharide synthesis gene(s) that encodes a 2-keto-3-deoxyoctulosonic acid epitope.** *Infect Immun* 1990, **58**:1558-1564.
9. Harvey HA, Jennings MP, Campbell CA, Williams R, Apicella MA: **Receptor-mediated endocytosis of Neisseria gonorrhoeae into primary human urethral epithelial cells: the role of the asialoglycoprotein receptor.** *Mol Microbiol* 2001, **42**:659-672.
10. Harvey HA, Swords WE, Apicella MA: **The mimicry of human glycolipids and glycosphingolipids by the lipooligosaccharides of pathogenic neisseria and haemophilus.** *J Autoimmun* 2001, **16**:257-262.
11. Harvey HA, Porat N, Campbell CA, Jennings M, Gibson BW, Phillips NJ, Apicella MA, Blake MS: **Gonococcal lipooligosaccharide is a ligand for the asialoglycoprotein receptor on human sperm.** *Mol Microbiol* 2000, **36**:1059-1070.
12. Phillips NJ, John CM, Reinders LG, Gibson BW, Apicella MA, Griffiss JM: **Structural models for the cell surface lipooligosaccharides of Neisseria gonorrhoeae and Haemophilus influenzae.** *Biomed Environ Mass Spectrom* 1990, **19**:731-745.
13. Johansen E, Gibson BW: **Functional Exploration of the *lsg* gene locus glycosyltransferases *lsgC, D, E, and F,* through chimeric expression in Eschericia coli K-12.** *Manuscript in Preparation* 2006.
14. Persing DH, Coler RN, Lacy MJ, Johnson DA, Baldridge JR, Hershberg RM, Reed SG: **Taking toll: lipid A mimetics as adjuvants and immunomodulators.** *Trends Microbiol* 2002, **10**:S32-37.

# APPENDIX

A. **Figure A-1:** Structures of bacterial LPS components. 108

B. **Table A-1:** Elemental compositions and masses of bacterial LPS components. 109

C. **Table A-2:** Partially methylated alditol acetates (PMAAs) derived from glycosidically linked sugar residues present in bacterial LPS. 110

D. **Table A3:** *Haemophilus influenzae lsg* locus gene accession numbers. 110

E. Improved protocol for isolation and purification of chimeric *E. coli* LPS. 111

F. Important concerns for chimeric *E. coli* LPS preparations. 112

G. **Figure A-2:** Negative-ion MALDI-TOF spectra from three preparations of O-deacylated LPS from *E. coli* K-12 transformant LOS-04 113

H. **Figure A-2:** Negative-ion MALDI-TOF spectra from three preparations of O-deacylated LPS from *E. coli* K-12 transformant LOS-05 114

## LPS Oligosaccharide Table

| Sugar | $M_r$ | In Chain: Nominal | Exact | Average |
|---|---|---|---|---|
| Hexose | 180<br>180.06339<br>$C_6H_{12}O_6$ | 162<br>$C_6H_{10}O_5$ | 162.052823 | 162.1424 |
| Heptose | 210<br>210.073955<br>$C_7H_{14}O_7$ | 192<br>$C_7H_{12}O_6$ | 192.06339 | 192.1687 |
| HexNAc | 221<br>$C_8H_{15}O_6N_1$ | 203<br>$C_8H_{13}O_5N_1$ | 203.079372 | 203.1950 |
| KDO | 238<br>$C_8H_{14}O_8$ | 220<br>$C_8H_{12}O_7$ | 220.058302 | 220.1791 |
| *anhdro*KDO | 220<br>$C_8H_{12}O_7$ | 202<br>$C_8H_{10}O_6$ | 202.04774 | 202.1638 |
| NeuAc | 309<br>$C_{11}H_{19}O_9N$ | 291<br>$C_{11}H_{17}O_8N$ | 291.095416 | 291.2579 |
| P | 98<br>$H_3PO_4$ | 80<br>$HPO_3$ | 79.966269 | 79.9799 |
| PEA | 141<br>$C_2H_8O_4N_1P_1$ | 123<br>$C_2H_6O_3N_1P_1$ | 123.0084678 | 123.0483 |
| O-DPLA | 952<br>$C_{40}H_{78}O_{19}N_2P_2$<br>(ave =953.0089) | 934<br>$C_{40}H_{76}O_{18}N_2P_2$ | 934.4567 | 934.9936 |
| Ac | 58<br>$C_2H_4O_2$ | 42<br>$C_2H_2O_1$ | 42.010565 | 42.0373 |
| $H_2O$ | | +18<br>$H_2O$ | 18.010565 | 18.0153 |

| Element | exact mass | average mass |
|---|---|---|
| C = 12 | 12.00000 | 12.011 |
| H = 1 | 1.007825 | 1.00794 |
| O = 16 | 15.9949146 | 15.9994 |
| N = 14 | 14.0030740 | 14.00674 |
| P = 31 | 30.973765 | 30.97376 |

**Table A-1.** Elemental compositions and masses of bacterial LPS components. O-DPLA represents *O*-deacylated diphosphorylated lipid A containing two N-linked β-hydroxy myristic acid chains.

**Table A-2. Partially methylated alditol acetates (PMAAs) derived from glycosidically linked sugar residues present in bacterial LPS.**

| In-Chain Residue | Corresponding PMAA |
|---|---|
| T-Gal | 1,5-Di-O-acetyl-1-deuterio-2,3,4,6-tetra-O-methyl-D-galactitol |
| 2-Glc | 1,2,5-Tri-O-acetyl-1-deuterio-3,4,6-tri-O-methyl-D-glucitol |
| 3-Gal | 1,3,5-Tri-O-acetyl-1-deuterio-2,4,6-tri-O-methyl-D-galactitol |
| 6-Glc | 1,5,6-Tri-O-acetyl-1-deuterio-2,3,4-tri-O-methyl-D-glucitol |
| 3,6-Glc | 1,3,5,6-Tetra-O-acetyl-1-deuterio-2,4-di-O-methyl-D-glucitol |
| T-Hep | 1,5,-Tetra-O-acetyl-2,3,4,6,7-tri-O-methyl-l-glycero-d-mannoheptitol |
| 3-Hep | 1,3,5-Tetra-O-acetyl-2,4,6,7-tri-O-methyl-l-glycero-d-mannoheptitol |
| 7-Hep | 1,5,7-Tetra-O-acetyl-2,3,4,6-tri-O-methyl-l-glycero-d-mannoheptitol |
| 3,7-Hep | 1,3,5,7-Tetra-O-acetyl-2,4,6-tri-O-methyl-l-glycero-d-mannoheptitol |
| T-GlcNAc | 1,5-Di-O-acetyl-2-(acetylmethylamino)-2-deoxy-1-deuterio-3,4,6-tri-O-methyl-D-glucitol |
| 4-GlcNAc | 1,4,5-Tri-O-acetyl-2-(acetylmethylamino)-2-deoxy-1-deuterio-3,6-di-O-methyl-D-glucitol |
| 3-GlcNAc | 1,3,5-Tri-O-acetyl-2-(acetylmethylamino)-2-deoxy-1-deuterio-4,6-di-O-methyl-D-glucitol |

**Table A-3: *Haemophilus influenzae lsg* locus gene accession numbers.**

| Gene | GenBank[A] Acession # | TIGR-CMR[B] Gene |
|---|---|---|
| *lsgA* | AAA24978.1 | HI1700 |
| *lsgB* | AAA24979.1 | HI1699 |
| *lsgC* | AAA24980.1 | HI1698 |
| *lsgD* | AAA24981.1 | HI1697 |
| *lsgE* | AAA24982.1 | HI1696 |
| *lsgF* | AAA24983.1 | HI1695 |
| *lsgG* | AAA24984.1 | HI1694 |

A: The *H. influenzae lsg* gene locus has the GenBank accession number of M98455 (http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?val=148931)

B: The Institute for Genomic Research – Comprehensive Microbial Resource (http://cmr.tigr.org/)

## IMPROVED PROTOCOL FOR ISOLATION AND PURIFICATION OF CHIMERIC *E. coli* LIPOPOLYSACCHARIDE.

**Day 01**

- Harvest ~80 ml liquid culture of bacteria. (plate preps do not work better for chimeric *E. coli*).
- Centrifuge suspension (10 min, 4°C, 7000 x *g*), discard supernatant.
- Wash bacteria with 50 ml PBS 1X, transfer into 2.0 ml Eppendorf tube.
- Centrifuge suspension (10 min, 4°C, 7000 x *g*), discard supernatant.
- Re-suspend bacterial pellet in 500 µl $H_2O$.
- Equilibrate both bacteria in $H_2O$ and bottle of JT Baker (VWR) liquefied phenol to 65°C 15 min.
- Mix 700 µl 65°C liquefied phenol with 700 µl (200 bacteria, 500 $H_2O$) bacteria.
- Incubate at 65°C, 45 min, vortex every 10 min.
- Cool tube in ice bath 15 min.
- Centrifuge (20 min, 4°C, 12000 x *g*) and separate phases.
- Save aqueous phase. Add 0.5 ml $H_2O$ to phenol phase and repeat previous three steps 1X.
- Concentrate pooled aqueuous phases in Savant speed-vac concentrator (on concentration more phenol that was dissolved in the $H_2O$ forms a second phase). When the sample is ~450 µl centrifuge to separate phases (20 min, 4°C, 12000 x *g*). Transfer aqueous phase into 2.0 ml Eppendorf tube.
- Add 1/10 volume 3M NaOAc pH 5.2 and 3 volumes EtOH: precipitate in -20 freezer overnight.

**Day 02**

- Centrifuge (20 min, 4°C, 12000 x *g*), separate pellet.
- Wash pellet with 1 ml EtOH.
- Dry down LOS pellet.
- Re-solubilize pellet into 450 µl 20 mM Tris-Cl (pH 8.0) + 2 mM $MgCl_2$
- Add 5 µl Roche DNAse (~10 µg /ml).
- Incubate at 37°C for 2 hours (mix occasionally or use orbital shaker).
- Add 1/10 volume 3M NaOAc pH 5.2 and 3 volumes EtOH: precipitate in -20 freezer overnight.

**Day 03**

- Pellet, dry, and re-suspend LOS in 2.0 mL $H_2O$.
- Ultracentrifuge 2 hours, 5°C, 120,000 x *g*.
- Solubilize pellet in 150 µl milli-Q water and transfer into pre-weighed 1.5 ml Eppendorf tube.
- Wash ultracentrifuge tube with 100 µl milli-Q water and add to the 1.5 ml Eppendorf tube.
- Dry 1.5 ml Eppendorf tube in Savant speed-vac concentrator ~2 hours and weigh sample.

**Protocol for O-deacylation of LOS:**

- Add 200-300 µl $H_2NNH_2$ to dry LOS pellet (~1 µg/µl – 10 µg/µl).
- Incubate at 37°C for 45 min with frequent vortexing.
- Cool tube in ice bath for 10 min.
- Slowly add 1000 µl of chilled acetone, then vortex (5X vol $H_2NNH_2$).
- Precipitate in -20 C freezer for 4 hours or overnight.
- Centrifuge (40 min, 4°C, 12000 x *g*), save pellet.
- Add 200 µl $H_2O$ to pellet.
- Dry down in speed vac and re-suspend pellet in 20 µl $H_2O$

## Important concerns for chimeric E. coli LPS preparations:

The standard LOS/LPS preparation when I joined this lab was to grow the bacteria on plates and then perform a phenol extraction to remove the LOS or LPS from the outer membrane of the bacteria. The protocol for this procedure was very similar to the protocol from day one of the improved protocol for chimeric LPS extraction and purification (above). This standard prep that was routinely used to prepare LOS from bacteria such as *H. ducreyi, H. influenzae, N. meningitides*, and *N. gonorrhoeae* with little difficulty, did not seem to work for the preparation of chimeric LPS from *E. coli.* After attempting the standard preparation a few times it became apparent that a new preparation be developed.

The first preparations attempted involved the use of lytic enzymes to degrade the three major competing biopolymers in a bacterial cell: DNA, RNA, and protein. After performing the phenol extraction of the LPS and precipitating overnight, LPS pellets were re-solubilized into buffer. Then a cocktail containing high concentrations of DNAse, RNAse, and Proteinase K was added to the LPS samples. This greatly increased the resolution of MALDI spectra of the enzyme-treated LPS samples (after O-deacylation, of course) relative to the standard prep, but large amounts of contamination remained.

The search for the optimum preparation of chimeric LPS from *E. coli* was expanded. I experimented with high and low concentrations of each of the three enzymes, as well as a number of different buffers (with and without SDS, etc.). I even experimented with ultracentrifugation techniques. Each parameter was tested rigorously against positive and negative controls. Only one parameter was varied at a time.

Approximately 50 LPS preparations later it was determined that treatment with low concentration of DNAse to remove the recombinant DNA, followed by ultracentrifugation at 120,000 x *g* for 2 hours resulted in the cleanest preparations of chimeric *E. coli* LPS performed to date in our laboratory. The negative-ion MALDI-TOF spectra of the O-deacylated LPS from *E. coli* transformed with the LOS-04 and LOS-05 vectors are featured in Figures A-2 and A-3, respectively. Each spectrum illustrates a qualitative determination of the efficacy of each method of preparing LPS. The data speaks for itself.

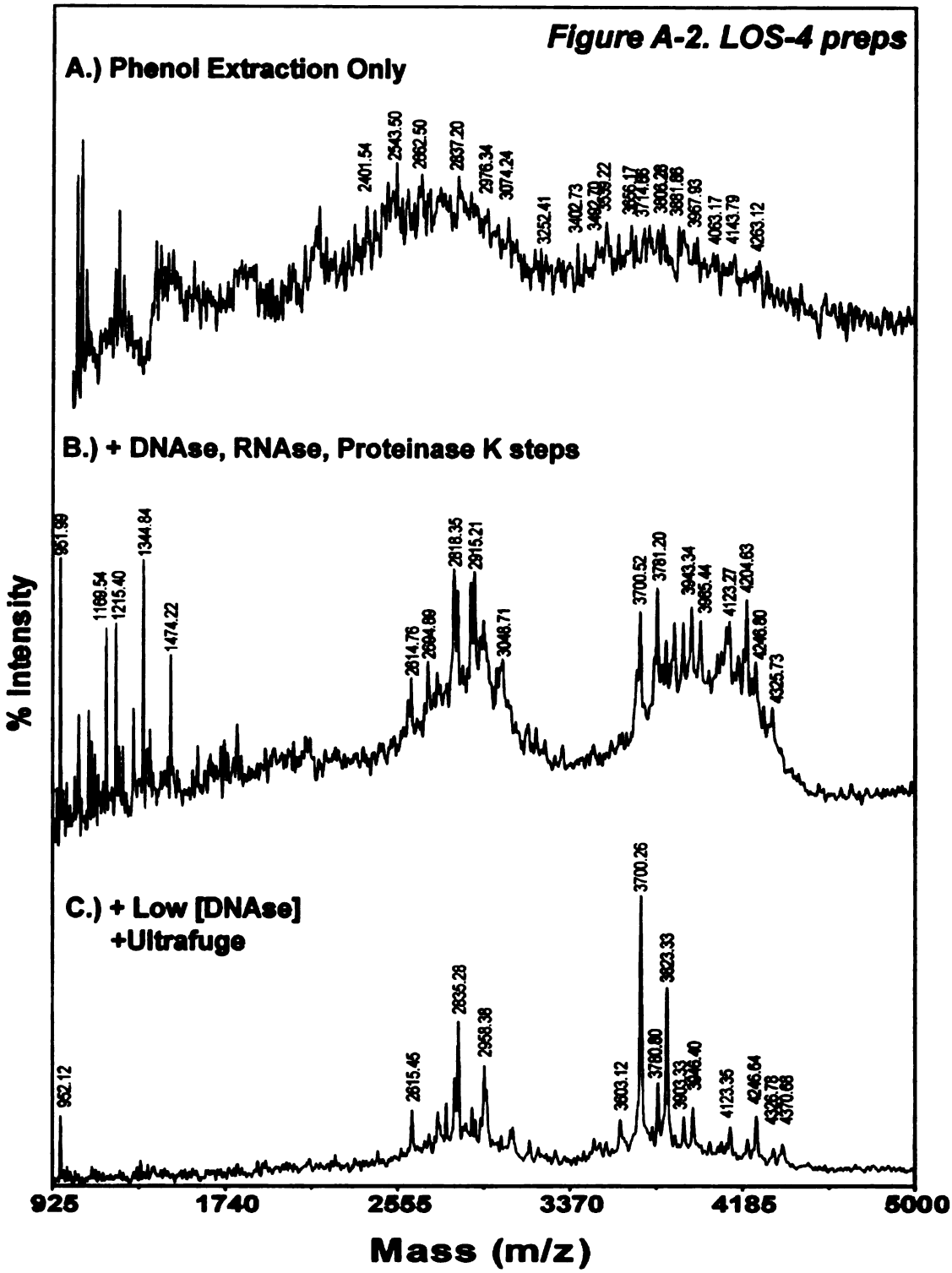



*Figure 4-2*: Negative ion MALDI-TOF spectra of the O-deacylated LPS from *E. coli* K-12 transformant LOS-04 from: A.) Phenol extraction only, B.) Phenol Extraction + treatment with 100 ug/ml DNAse, 500 ugml RNAse, and 200 ug/ml Proteinase K, and C.) Phenol extraction and treatment with 10 ug/ml DNAse + Ultracentrifugation @ 120,000 x *g*. for 2 hr.

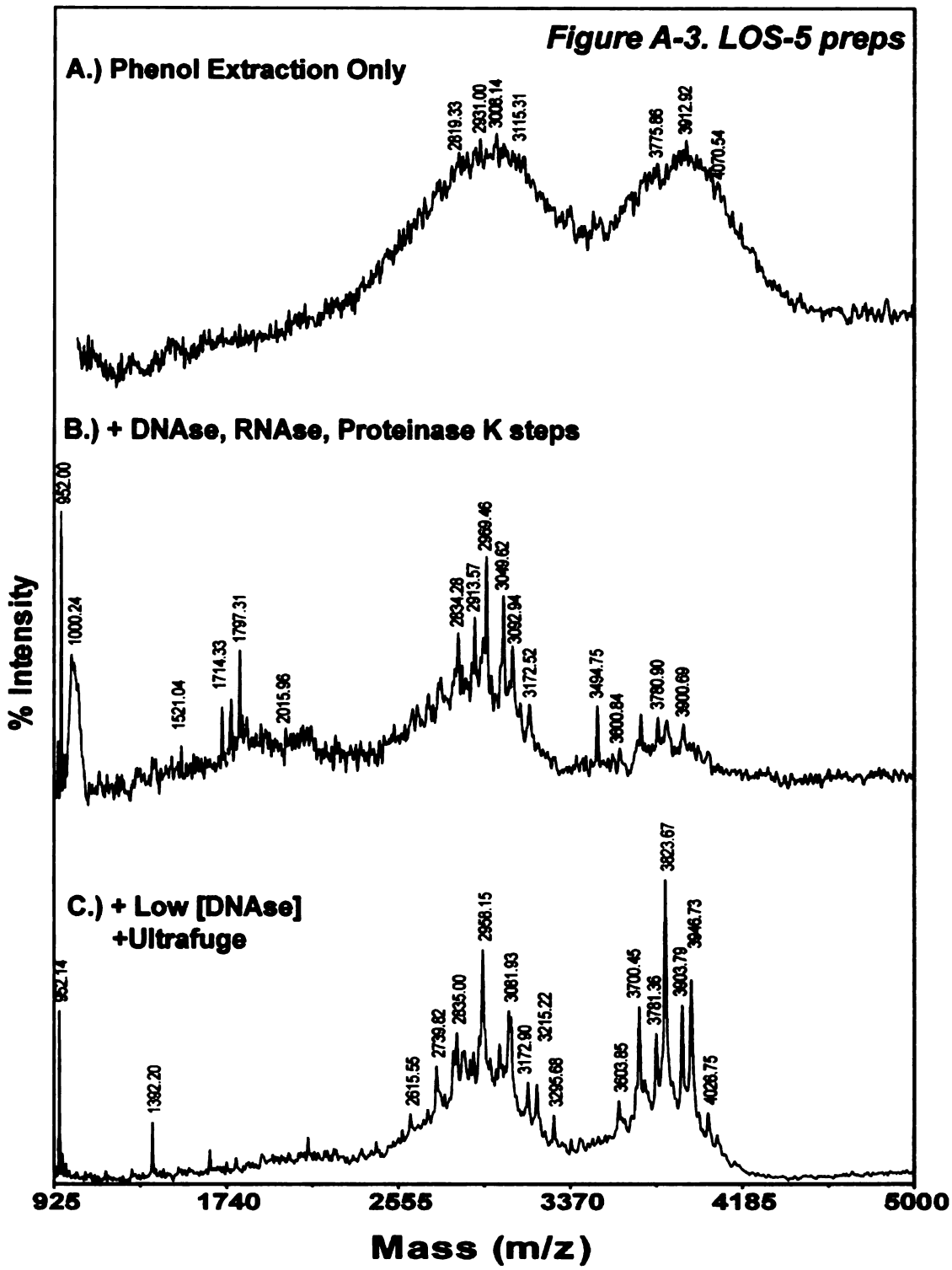



***Figure 4-2***: Negative ion MALDI-TOF spectra of the O-deacylated LPS from *E. coli* K-12 transformant LOS-05 from: A.) Phenol extraction only, B.) Phenol Extraction + treatment with 100 ug/ml DNAse, 500 ugml RNAse, and 200 ug/ml Proteinase K, and C.) Phenol extraction and treatment with 10 ug/ml DNAse + Ultracentrifugation @ 120,000 x *g*. for 2 hr.

UCSF LIBRARY

7540910
3 1378 00754 0910
UC San Francisco LIBRARY