DIFFERENTIAL GENE EXPRESSION AND COLANIC ACID GENE EFFECTS ON BIOFILM FORMATION

THESIS

Presented to the Graduate Council
of Texas State University-San Marcos
in Partial Fulfillment
of the Requirements
for the Degree

Master of SCIENCE

by

Christa Lynn Bates, B.S.

San Marcos, Texas
December 2004

ACKNOWLEDGEMENTS

Special thanks are set forth to my husband, Kris French, for his encouragement to further my education along with support throughout. Special thanks to my parents for assistance and moral support throughout my education. Extensive appreciation goes to Mary Barnes and Dr. Robert McLean for assistance in learning new techniques and lending much needed support throughout my undergraduate and graduate careers. Thanks to my committee members Dr. Mike Forstner, Dr. Gary Aron, and Dr. Deborah Siegele for their guidance and assistance in completing my thesis. Thanks to Kerry Fuson for providing substantial data and preliminary studies which dealt with gene expression in biofilms. I would also like to thank all members of the McLean lab for their assistance and support. A further special thanks is endowed to Dr. Robert McLean for allowing me to accomplish my goals while working in his lab along with extensive patience and understanding of adverse events experienced during my graduate career.

This manuscript was submitted on December 1, 2004.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

ABSTRACT

DIFFERENTIAL GENE EXPRESSION AND COLANIC ACID GENE AFFECTS ON BIOFILM FORMATION

by

Christa Lynn Bates, B.S.

Texas State University-San Marcos
December 2004

SUPERVISING PROFESSOR: ROBERT J.C. MCLEAN

Biofilm gene expression is an emerging field of study within bacterial genomics. Investigators strive to understand changes in phenotype, morphology, and metabolism in response to selective pressures within various environments. By understanding the genetic switches activated as a result of phenotype change we may more readily control the growth of non-beneficial bacterial communities. This study analyzed general differential expression due to the maturity of the biofilm community through statistical analyses of microarray data from four and six day biofilm cultures. Molecular sequence analyses were performed to demonstrate genomic affects of insertion or deletion mutations. Lastly, characterization of insertion mutations of *wzc* with respect to biofilm formation was performed to assess the role of colanic acid in newly formed biofilms. A

significant difference was found between gene expression and maturity using the ANOVA. The post mortem analyses of biofilm cultures found significant changes in genes with functions in translation, ribosomal structure and biogenesis, transcription, DNA replication, recombination and repair, cell division and chromosome partitioning, cell envelope biogenesis and outer membrane synthesis, energy production and conversion, carbohydrate transport and metabolism, coenzyme metabolism, lipid metabolism, and secondary metabolite biosynthesis and transport catabolism. These finding suggest a metabolic shift as a result of maturation between four and six day biofilms. A major oversight in current research is pre-mutagenesis sequence analyses to determine the potential genomic expression affects. Molecular sequence analyses were performed to demonstrate methods for pre-mutagenesis sequence characterization. Sequences of up-, down-, and neutrally regulated genes were analyzed along with surrounding sequence up and downstream. Sequences were analyzed in six frames for amino acid sequence, potential overlapping open reading frames (ORFs), protein homology, conserved domain presence, RNA secondary structure, and protein locality. Up-regulated b1995 is potentially involved in outer membrane permeability control, is possibly a trans-acting factor, and is most likely located in the periplasmic area between the inner and outer membranes of Gram-negative bacteria. Neutrally transcribed, b1394 is potentially involved in operons controlling the switch between aerobic and anaerobic growth with a protein localization of cytoplasmic nature. Down-regulated gene, b2852 is suspected to be involved in some type of membrane permeability control, however these results were not clear and the protein localization could not be predicted. These predictions will be used in future work as a basis for analyzing phenotypical and

morphological changes post-mutagenesis. Specific analyses were performed to assess the role of colanic acid, exopolysaccharide, synthesis on early biofilm formation. *Escherichia coli* JM83 and *Escherichia coli* JM83wzc::KmR were transformed with plasmids which expressed functional and single amino acid mutations for *wzc*. This gene is suggested to prevent colanic acid production when mutated. Colanic acid production, namely fucose a non-dialyzable methyl-pentose, was quantified by growth of all strains in nutrient rich and minimal medias followed by centrifuge extraction from broth, dialysis, lyophilization, concentration, and standardized against L-fucose. Biofilm production was quantified via methods described by George O'Toole and Roberto Kolter. Amounts of colanic acid produced did not correlate to the functionality of *wzc*, which suggests its lack of influence on exopolysaccharide synthesis within the first 48 hours of growth. It was also found that functionality of *wzc* does not correlate to biofilm formation. Moreover, mutations in phosphorylation and catalytic sites for colanic acid production may promote biofilm formation. The information obtained herein can be further applied to targeting specific genes and gene products for promoting or discouraging biofilm formation.

CHAPTER I

STATISTICAL EVALUATION OF *ESCHERICHIA COLI* GENE EXPRESSION CHANGES IN BIOFILMS DURING COMMUNITY MATURITY

INTRODUCTION

Bacteria are capable of expressing two phenotypes, however biofilms are the most common growth form in the environment. This phenotype allows growth under adverse conditions and resistance to antibiotics or other toxic substances due to morphological changes (KROOS and MADDOCK 2003). Morphological changes that result as a product of environmental changes are of great interest in medical, industrial, and environmental situations. By understanding the genetic switches that occur as a result of phenotype change we may more readily control the growth of non-beneficial bacterial communities.

Biofilm gene expression is an emerging field of study within bacterial physiology. Investigators strive to understand changes in phenotype, morphology, and metabolism (COVERT and PALSSON 2002) in response to selective pressures within various environments. Gene expression profiling allows researchers to assess the concentration and hence, expression of specific gene transcripts during different growth conditions. DNA microarrays have emerged as the common means for performing these analyses. This method uses messenger RNA (mRNA) transcripts harvested from the growing cells, which are then hybridized to specific probes (sequences) on a membrane. The transcripts

can be fluorescently labeled and as they hybridize to the probe the amount of fluorescence often increases depending on the number of transcripts bound. The relative number of transcripts present often represents the level of the gene product produced, hence the amount of protein needed by the metabolizing cell (GRUNBERG-MANAGO 1996). Investigators currently believe that different levels of production of different genes depends on the level of starvation of the community (TAO *et al.* 1999). Current studies analyze differences in gene expression between nutrient rich and poor media (TAO *et al.* 1999; WEI *et al.* 2001) to determine the genome-wide expression changes between different growth conditions.

With the advent of the entire *Escherichia coli* genome sequence (BLATTNER *et al.* 1997) and manufactured microarrays, genomic expression studies have climbed to a new level of knowledge allowing investigators to determine genes involved in biofilm formation (BALZER and MCLEAN 2002; FIROVED and DERETIC 2003; FUSON 2002; O'TOOLE and KOLTER 1998; PRIGENT-COMBARET *et al.* 2001; PRIGENT-COMBARET *et al.* 1999; TANI *et al.* 2002). These genes encode activities ranging from viability to growth characteristic changes. Microarray analyses allow for identification of potential mutation targets for promotion or prevention of biofilm formation, based on their expression rates when growth is selected for a specific morphology and allows analysis of potential growth interactions due to environmental conditions.

Specifically, expression levels in *E. coli* K-12 have been studied as the ratio of biofilm over planktonic cDNA transcripts binding antisense oligonucleotide microarrays to discern genes used extensively in biofilm and planktonic cells (SCHEMBRI *et al.* 2003). This information facilitates focusing on genes that are transcribed often in biofilm cells

and hence potential protein, receptor, or enzyme targets for disaggregating attached communities of bacteria. Through the understanding of potential up-regulated biofilm genes, investigators can identify specific gene products that may be involved in attachment, microcolony formation, and resistance (STEWART and COSTERTON 2001) among other survival factors for biofilms. Recognition of the categories of genes that are expressed in different environments or physiological states and operon structures (SABATTI *et al.* 2002) of differential regulation are now possible.

Current knowledge of differential gene expression in biofilms versus their planktonic counter parts encompasses transcription of genes in one physiological state and not in others, (ARNOLD *et al.* 2001; AVISON *et al.* 2001; JACKSON *et al.* 2002) and cell structure control such as seen with *csgD* in *Escherichia coli* biofilms (PRIGENT-COMBARET *et al.* 2001). However, there is not much research on the genomic level changes in expression, which focus on general gene types that undergo differential expression during changes in the maturity or environmental conditions of the community.

To fully understand how expression occurs within bacterial cells one must take into account the changes in metabolism and other factors due to small changes in maturity, meaning expression levels of a particular gene at one phase in biofilm formation may not be the same in an earlier or later stage. Also of interest are the types of genetically controlled processes that undergo significant changes throughout the phases of biofilm growth. The aim of this study is to analyze general differential expression due to the maturity of the biofilm community through statistical analyses of microarray data from four and six day biofilm cultures.

MATERIALS AND METHODS

Microarray gene expression data were obtained from previous work in our lab that is currently in preparation for publication (FUSON 2002). The data were expressed as the ratio of bound transcript for biofilm over planktonic rRNA. The arrays were based on two replicates of four and six day *Escherichia coli* biofilms grown in MOPS minimal defined medium as described in previous literature (NEIDHARDT *et al.* 1974) with serine (1mg/mL) as a limited carbon source and amino acids required by acid auxotrophic strains (Ile, Arg, Gly, His, Leu, Met, Phe, Val, Thr, each at 4mg/mL). Eighteen gene function categories are currently present in *E. coli* and were used to determine the major areas of change during biofilm maturation (NCBI COGs Functional Annotation, http://www.ncbi.nlm.nih.gov/COG/old/palox.cgi?fun=all). Subsets of seven genes from each gene function category in *E. coli* were randomly chosen for analysis. See Figure 1 for an illustration of the genes and gene function categories used along with the experimental design. All analyses were performed using S-Plus 6.1 (Insightful Corporation). A three-factor balanced analysis of variance with 7 randomly chosen genes (from the respective gene function category) nested in gene function, gene function fixed between the four and six day cultures and independent of maturity, and maturity fixed and independent of gene function. This analysis was used to assess the amount of variation between biofilm maturity, gene function category, and gene within gene function category based on expression levels. Data were natural log transformed to meet the assumptions of normality and homoscedasticity. Further two-factor analyses of variance were performed for each gene function category to conduct post mortem

analyses. These analyses used maturity and gene function as factors with the expression rate for response. Data were also log transformed to meet assumptions.

RESULTS

Results of the ANOVA illustrated dependence in gene expression levels between maturity and gene within gene function category (Table 1). The interaction between maturity and gene within gene function nullified the significance of any other variance within the analyses. The significance of this interaction ($p=0.005$) suggested differences in expression between the two maturity levels of the biofilms with respect to gene and gene function.

Based on these results, additional post mortem analyses were performed using Tukey's HSD multiple comparison procedure. The multiple comparison procedure analyzed the difference in expression of gene function categories for each maturity level. After eighteen Tukey's were performed some gene function categories failed to differ in their expression rates while others illustrated significantly different expression levels (Table 2). It was concluded that significant gene expression changes occur with respect to genes responsible for translation, ribosome structure synthesis and biogenesis, transcription, DNA replication, DNA recombination, DNA repair, cell envelope biogenesis, outer membrane biogenesis, energy production and conversion, carbohydrate transport and metabolism, coenzyme metabolism, lipid metabolism, secondary metabolite biosynthesis, and transport catabolism maturation (NCBI COGs Functional Annotation, http://www.ncbi.nlm.nih.gov/COG/old/palox.cgi?fun=all). Also, the expression level means for six-day maturity were greater in categories of gene function that proved significantly different between four and six-day cultures. No significant differences were

found in expression levels of genes responsible for cell division, chromosome partitioning, post-translational modification, protein turnover, chaperones, cell motility, secretion, inorganic ion transport and metabolism, signal transduction mechanisms, amino acid transport and metabolism, nucleotide transport and metabolism, general function genes that are only predicted, and functionally unknown genes. In reviewing the means of expression ratios for four and six day cultures, six day cultures were found to illustrate greater expression levels in all significant gene function categories.

DISCUSSION

It was determined that when biofilm communities differ in maturity, their gene expression profiles reflect these differences. Interaction between maturity and gene within gene function suggests a significant change in gene expression during the maturation process. Post mortem analyses using Tukey's HSD multiple comparison procedures found the gene function categories that significantly differed between the maturity levels with greater means at six days. These findings can be attributed to establishment of the biofilm community resulting in faster growth rates or possibly preparation for dispersion of cells from the biofilm. However, in preparation for dispersion one should observe significant differences in cell motility and secretion genes, which were not the case at six days of maturity. Therefore, one could suggest the community is in very early stages of full maturity. With the non-significance of cell division and chromosome partitioning genes one could argue that the up-regulation of other genes involved in new cell synthesis is actually a factor of exopolysaccharide synthesis, which protects mature biofilms from adverse environments (DANESE *et al.* 2000b). Several genes have been found to play a role in expopolysaccharide and

lipopolysaccharide production and regulation, and when up-regulated, lead to biofilm formation (BALZER and MCLEAN 2002; EBEL and TREMPY 1999; GOTTESMAN *et al.* 1985; GUPTE *et al.* 1997; WEHLAND and BERNHARD 2000). Another hypothesis could include changed metabolic properties and protein localization due to the maturing community (COVERT and PALSSON 2002; KUMARI *et al.* 2000). This hypothesis seems most plausible due to the fact that as more cells are present in an area, the amount of waste increase and nutrients decrease. The metabolic pathways which occur in the cells are dependent on the trans-acting factors produced by earlier pathways (COVERT and PALSSON 2002). This theory is supported by increases in energy production and conversion, carbohydrate transport and metabolism, coenzyme metabolism, secondary metabolite biosynthesis, and transport catabolism. For these processes to occur the cell must also transcribe and translate differential genes hence supporting the increase in translation, ribosomal structure, DNA replication, recombination and repair, and transcription gene level controls. The increase in cell envelope biogenesis and outer membrane synthesis could be explained by changes in the cell membranes to compensate for metabolic switches or by the production of polysaccharides implemented in biofilm formation (DANESE *et al.* 2000a).

Previous studies have focused on the morphological changes as biofilms mature (TOLKER-NIELSEN *et al.* 2000) however, these studies do not represent the changes in the transcriptome that occur as the communities mature. Currently, investigators have only described gene expression differences between planktonic and biofilm (SCHEMBRI *et al.* 2003). By understanding gene expression changes over time, one can identify potential targets for biofilm disaggregation at any stage of maturity and understand when the genes

for protein or enzyme products that cause the most adverse effects to ecological systems, health, and industrial processes are expressed.

LITERATURE CITED

ARNOLD, C. N., J. MCELHANON, A. LEE, R. LEONHART and D. A. SIEGELE, 2001 Global Analysis of Escherichia coli Gene Expression during the Acetate-Induced Acid Tolerance Response. J. Bacteriol. **183:** 2178-2186.

AVISON, M. B., R. E. HORTON, T. R. WALSH and P. M. BENNETT, 2001 Escherichia coli CreBC Is a Global Regulator of Gene Expression That Responds to Growth in Minimal Media. J. Biol. Chem. **276:** 26955-26961.

BALZER, G. J., and R. J. C. MCLEAN, 2002 The stringent response genes *relA* and *spoT* are important for *Escherichia coli* biofilms under slow-growth conditions. Canadian Journal of Microbiology **48:** 675-680.

BLATTNER, F. R., G. PLUNKETT III, P. L. BLOCH, N. T. PERNA, V. BURLAND *et al.*, 1997 The complete genome sequence of *Escherichia coli* K-12. Science **277:** 1453-1462.

COVERT, M. W., and B. O. PALSSON, 2002 Transcriptional regulation in constraints-based metabolic models of *Escherichia coli.* Journal of Biological Chemistry **277:** 28058-28064.

DANESE, P. N., L. A. PRATT and R. KOLTER, 2000a Exopolysaccharide production is required for development of *Escherichia coli* K-12 biofilm architecture. Journal of Bacteriology **182:** 3593-3596.

DANESE, P. N., L. A. PRATT and R. KOLTER, 2000b Exopolysaccharide Production Is Required for Development of Escherichia coli K-12 Biofilm Architecture. J. Bacteriol. **182:** 3593-3596.

EBEL, W., and J. E. TREMPY, 1999 Escherichia coli RcsA, a Positive Activator of Colanic Acid Capsular Polysaccharide Synthesis, Functions To Activate Its Own Expression. J. Bacteriol. **181:** 577-584.

FIROVED, A. M., and V. DERETIC, 2003 Microarray Analysis of Global Gene Expression in Mucoid Pseudomonas aeruginosa. J. Bacteriol. **185:** 1071-1081.

FUSON, K. L., 2002 Biofilm-induced gene expression in chemostat grown *Escherichia coli* as determined by a gene array, pp. 60 in *Biology.* Texas State University, San Marcos, Texas.

GOTTESMAN, S., P. TRISLER and A. TORRES-CABASSA, 1985 Regulation of capsular polysaccharide synthesis in *Escherichia coli* K-12: characterization of three regulatory genes. J. Bacteriol. **162:** 1111-1119.

GRUNBERG-MANAGO, M., 1996 Regulation of the expression of aminoacyl-tRNA synthetases and translation factors, pp. 1432-1457 in *Escherichia coli and Salmonella: cellular and molecular biology*, edited by F. C. NEIDHARDT, R. CURTISS III, J. L. INGRAHAM, E. C. C. LIN, K. B. LOW *et al.* ASM Press, Washington, DC.

GUPTE, G., C. WOODWARD and V. STOUT, 1997 Isolation and characterization of rcsB mutations that affect colanic acid capsule synthesis in Escherichia coli K-12. J. Bacteriol. **179:** 4328-4335.

JACKSON, D. W., K. SUZUKI, L. OAKFORD, J. W. SIMECKA, M. E. HART *et al.*, 2002 Biofilm Formation and Dispersal under the Influence of the Global Regulator CsrA of Escherichia coli. J. Bacteriol. **184:** 290-301.

KROOS, L., and J. R. MADDOCK, 2003 Prokaryotic Development: Emerging Insights. J. Bacteriol. **185:** 1128-1146.

KUMARI, S., C. M. BEATTY, D. F. BROWNING, S. J. W. BUSBY, E. J. SIMEL *et al.*, 2000 Regulation of Acetyl Coenzyme A Synthetase in Escherichia coli. J. Bacteriol. **182:** 4173-4179.

NEIDHARDT, F. C., P. L. BLOCH and D. F. SMITH, 1974 Culture medium for enterobacteria. J. Bacteriol. **119:** 736-747.

O'TOOLE, G. A., and R. KOLTER, 1998 Flagellar and twitching motility are necessary for *Pseudomonas aeruginosa* biofilm development. Mol. Microbiol. **30:** 295-304.

PRIGENT-COMBARET, C., E. BROMBACHER, O. VIDAL, A. AMBERT, P. LEJEUNE *et al.*, 2001 Complex Regulatory Network Controls Initial Adhesion and Biofilm Formation in Escherichia coli via Regulation of the csgD Gene. J. Bacteriol. **183:** 7213-7223.

PRIGENT-COMBARET, C., O. VIDAL, C. DOREL and P. LEJEUNE, 1999 Abiotic Surface Sensing and Biofilm-Dependent Regulation of Gene Expression in Escherichia coli. J. Bacteriol. **181:** 5993-6002.

SABATTI, C., L. ROHLIN, M. K. OH and J. C. LIAO, 2002 Coexpression pattern from DNA microarray experiments as a tool for operon prediction. Nucleic Acids Research **30:** 2886-2893.

SCHEMBRI, M. A., K. KJAERGAARD and P. KLEMM, 2003 Global gene expression in *Escherichia coli* biofilms. Mol. Microbiol. **48:** 253-267.

STEWART, P. S., and J. W. COSTERTON, 2001 Antibiotic resistance of bacteria in biofilms. Lancet **358:** 135-138.

TANI, T. H., A. KHODURSKY, R. M. BLUMENTHAL, P. O. BROWN and R. G. MATTHEWS, 2002 Adaptation to famine: A family of stationary-phase genes revealed by microarray analysis. PNAS **99:** 13471-13476.

TAO, H., C. BAUSCH, C. RICHMOND, F. R. BLATTNER and T. CONWAY, 1999 Functional genomics: expression analysis of *Escherichia coli* growing on minimal and rich media. J. Bacteriol. **181:** 6425-6440.

TOLKER-NIELSEN, T., U. C. BRINCH, P. C. RAGAS, J. B. ANDERSEN, C. S. JACOBSEN *et al.*, 2000 Development and dynamics of *Pseudomonas* sp. biofilms. J. Bacteriol. **182:** 6482-6489.

WEHLAND, M., and F. BERNHARD, 2000 The RcsAB Box. Characterization of a new operator essential for the regulation of expolysaccharide biosynthesis in enteric bacteria. J. Biol. Chem. **275:** 7013-7020.

WEI, Y., J.-M. LEE, C. RICHMOND, F. R. BLATTNER, J. A. RAFALSKI *et al.*, 2001 High-Density Microarray-Mediated Gene Expression Profiling of Escherichia coli. J. Bacteriol. **183:** 545-556.

CHAPTER II

MOLECULAR SEQUENCE CHARACTERIZATION OF NOVEL GENES INVOLVED IN *ESCHERICHIA COLI* PHENOTYPE EXPRESSION

INTRODUCTION

A major oversight within the biofilm gene expression field is pre-mutagenesis molecular sequence analysis. These analyses require amino acid sequence comparisons of not only the gene of interest (GOI) but also surrounding upstream and downstream regions. Pre-mutagensis analysis allows for determination of potential interactions with other genes or gene products. When knock-out mutagenesis is employed one must be aware of the potential effects to the remainder of the genome with respect to overlapping open reading frames (ORFs), cis-acting elements, and trans-acting factors that may affect the determination of gene function. These types of genes have been characterized in *E. coli* such as the BglG protein, which controls RNA polymerase via a trans interaction for the bgl operon (Nussbaum-Shochat and Amster-Choder 1999). Assessing the potentiality of trans or cis interactions from gene products is important due to the genome-wide control of expression. Trans or cis reactions are based on the interaction of the element or factor with other genes, which may be necessary for varied metabolic processes, and hence make the confidence of mutation results undeterminable. The aim of this study is

to convey the methods that should be employed when considering a gene for knock-out mutagenesis and predict the potential genomic and physiological outcome of gene deletions for b1995, b2852, and b1394 which are up-regulated, neutral, and down-regulated in *E. coli* biofilms.

MATERIALS AND METHODS

Sequence analysis was based on the *Escherichia coli* K-12 MG1655 version M52 currently published through University of Wisconsin at *http://www.genome.wisc.edu/sequencing.htm.*

Sequences were extracted for genes b1995, b2852 and b1394 which were found to be up-regulated, neutrally transcribed and down-regulated, respectively, based on four and six day biofilm and planktonic cultures from a MOPS minimal medium chemostat microarray analysis (Fuson 2002). Sequences were extracted for the gene only, the gene plus one, two and three genes up and down stream, and the gene plus 5,000 bases up and down stream of the N and C terminals for the genes of interest. BioEdit was used as a sequence analysis program available publicly via *http://www.mbio.ncsu.edu/BioEdit/bioedit.html.* The predicted amino acid sequences were translated directly in BioEdit and then used further as the basis for molecular analysis.

All sequences were analyzed in all six frames for amino acid sequence, potential overlapping open reading frames, and conserved protein domains. Potential overlapping reading frames were assessed using NCBI's ORF Finder, which analyzed the six frames for open reading frames based on the nucleotide sequence for stop and start codons.

Conserved domain and protein similarity analyses were performed using NCBI's Protein-Protein BLASTP.

Frames considered to express the gene of interest open reading frame and substantial known protein similarity were further analyzed for RNA secondary structure. Sequences were first translated to amino acid sequence. GeneBee was used to predict RNA secondary structure based on potential folding patterns. SubLoc 1.0 analyzed the potential localization of the protein products from the genes. Based on these analyses potential effects of knockout mutations were assessed.

RESULTS

Amino acid sequences are shown in Table 1 for each gene sequence in the frame that proved most probable for producing the protein product and was similar to other proteins. Overlapping reading frames were present for all genes however, b1995 contained no potential ORFs in frame with the start codon. One potential ORF in the –3 frame (four bases upstream of the first nucleotide of the start codon would start the codon frames) was found between nucleotides 836 and 970 of b2852. The potential internal ORF overlap for b2852 produced significant amino acid sequence similarity with o458 (b2852) (100%), a putative invasion protein from K-12 (100%), and a putative invasion protein from O157:H7 (97%). For b1394 there was also one overlapping ORF in frame –3 from bases 407 to 619. However, these alignments were based on only 47 amino acids, which was not a significant representation of the subject sequence. For b1394 internal ORF similarity was not significant.

NCBI Protein-Protein BLAST produced significant results for b1995 and b2852, however there were no significant (>=90% identity) similarities for b1394. Sequences similarity for b1995 at 100% was seen for alignment against its own sequence, a hemin receptor precursor for *E. coli* K-12, and a putative outer membrane receptor for iron or colicin, an unknown protein, and a hypothetical protein all present in *E. coli* O157:H7. Similarity was also seen with a putative outer membrane iron or colicin receptor for *E. coli* CFT073. Significant alignments for b2852 were a putative invasion protein in K-12 with 97% identity, open reading frame o458 with 96%, and a putative invasion protein from O157:H7 at 94% identity.

Conserved domain alignments, considered 99% alignment or better, for b1995 predicted homology in frame 1 to CadC at 100% amino acid sequence identity and Shiga-like toxin (SLT) also at 100%. These alignments were observed slightly downstream of b1995 within the 5,000 nt up and downstream sequence. Transcription regulatory protein C terminal was aligned (100%) within the C terminal region of b2852 as was CadC also at 100% aligned. Significant conserved domain alignments for b1394 illustrated similarity of CaiD (98.8%) within the gene. For regions surrounding the gene, significant alignments downstream were represented with PaaJ (100.0%), Thiolase (100.0%), and Thiolase_C (99.2%). Upstream identity was represented by PaaA and PaaC (99.6%), COG3396 (100%), PaaD (100%), and pfam01883 DUF59 (100%). Locations of conserved domain significant protein alignments are shown in Table 2 and Table 3, respectively.

RNA secondary structure and secondary structure stability were analyzed using GeneBee. Stability for b1995 RNA secondary structure was predicted at –54.3 kcal/mol

with stem stability ranging from –6.2 kcal/mol to –15.0 kcal/mol. RNA secondary structure stability for b2852 was predicted as –155.8 kcal/mol with stem stability ranging from –6.2 kcal/mol to –16.3 kcal/mol. Finally, stability for b1394 RNA was predicted at –163.4 kcal/mol with stem stability ranging from –6.1 to –17.5 kcal/mol. RNA secondary structures are shown in Figure 1.

SubLoc v1.0 was used to predict protein localization. Potential localization of b1995 was predicted as periplasmic with a reliability index of 5 (a maximum of 10 possible) and an expected accuracy of 96%. Localization for b2852 was predicted as cytoplasmic with a reliability index of 9 and expected accuracy of 97%. Finally, b1394 was predicted to localize within the cytoplasm with a reliability index of only 1 and an expected accuracy of only 63%.

DISCUSSION

Differential gene expression allows us to determine the effects of differential growth conditions on bacterial transcription products, which can be analyzed further, at the genomic level. By understanding the effects of differential gene expression and differential gene category expression at a genomic level further analyses for phenotype, viability, and metabolism can be predicted. The genomic implications require a knowledge of protein localization, motif conservation, and gene family interaction. To assess these questions we evaluated up-, neutral-, and down-regulated genes using amino acid sequence for overlapping ORF potentiality, protein similarity, conserved domain similarity, RNA secondary structure, and potential protein localization. These data were then used to predict the potential results of a knockout mutation.

Based on the NCBI ORF Finder results the removal of b1995 would not affect any overlapping ORFs hence, the potential for removing other genes is not a concern. Due to the short length of the ORF sequence overlapping b2852 we also conclude there will not be any concern for removing sequences of other potential genes. Overlapping ORFs for b1394 did not produce significant amino acid sequence similarity to any known or uncharacterized genes, therefore no potential coding regions other than the gene of interest will be affected.

The protein-protein BLAST results show the protein product of b1995 is homologous with an iron or hemin outer membrane receptor. Being that this gene is up-regulated in biofilm phenotypes, it is suggested that bacteria may need to evolve different mechanisms for iron or metabolite uptake as compared to their planktonic counterparts, which can relocate to areas of increased metabolites or may in fact need less of the metabolite. This conclusion is made due to the inherent structure of biofilms and the suggestion that they undergo co-metabolism (MOLLER *et al.* 1998) within microcolonies and communities hence needing enough metabolite to supply and ensure the viability of the entire aggregate. Homology was seen between b2852 and a putative invasion protein present in *E. coli* K-12, open reading frame o458, and a putative invasion protein from O157:H7. This gene was considered a neutrally-regulated coding region in biofilms and invasion or infection proteins are most likely expressed in both phenotypes, supported by urinary tract infections due to biofilms and upper-gastrointestinal infections due to planktonics. Based on the insignificant results of the protein-protein BLAST, a protein product of b1394 was not predicted.

Conserved domain analyses identified a high degree of sequence similarity between a sequence just upstream of b1995 with CadC and Shiga-like toxin (SLT). CadC is a regulator of the CadAB operon, which is responsible for acid tolerance based on reducing the outer-membrane porosity of Gram-negative bacteria (Samartzidou and Delcour 1999). The sequence of this motif suggests an upstream control of an outer-membrane porosity factor. It has been suggested that the inherent antibiotic resistance of biofilms relies on reduced permeability to β-lactams, which act in the periplasmic space between the outer and inner-membranes of Gram-negative bacteria (Nikaido 1996), which suggests the reason for up-regulation of genes that control membrane permeability in biofilms. The conservation with SLT, in particular with the motif component OmpC, a receptor for CpxR-P suggests the same conclusion. CpxR-P regulates expression of gene products also involved in reduced membrane porosity, relieving envelope protein distress, regulated biofilm formation, motility and chemotaxis, host cell invasion, and virulence (De Wulf *et al.* 2002). Therefore, the removal of this portion of the outer membrane control motif may potentially decrease the resistance of biofilms to sheer forces and toxic substances. The RNA secondary structure, illustrated in Fig. 1-A, shows many A-T rich regions, which are common to CpxR-P receptors and other periplasmic and membrane proteins (De Wulf *et al.* 2002). The analysis of potential protein localization also supported this hypothesis with a 96% assurance that this coding region resulted in a periplasmic protein. This gene product may also be a trans-acting factor as seen with CadC, hence one should take into account the potential genomic expression effects.

Conservation of sequence with b1394, the neutrally expressed gene of interest, was represented by CaiD within the gene, PaaJ, thiolase and thiolase_C downstream,

PaaA, PaaC, COG3396, PaaD, and pfam01883 DUF59 upstream. CaiD, a component of the *caiTABCDE* operon, is involved in carnitine pathway, and more specifically, produces enoyl hydratase-isomerase during anaerobic growth(BUCHET *et al.* 1998). The ability to grow under anaerobic conditions is important for both planktonic and biofilm phenotypes. Downstream similarity was represented by PaaJ, a constituent of the phenylacetic acid (PA) operon in *E. coli* K-12 (FERRANDEZ *et al.* 1998). This operon is useful as PA is a common source of carbon and energy during aerobic growth (FERRANDEZ *et al.* 1998). This suggests a gene switch for CaiD and PA operon during anaerobic and aerobic growth, respectively; hence the gene of interest may be an intermediate in this switch or a portion of one of these operons. Thiolase, located downstream, is also associated with aromatic compound degradation for utilization of carbon sources (DIAZ *et al.* 2001).

The down-regulated gene of interest, b2852 illustrated internal similarity with CadC and Transcription regulatory protein (C terminal). This similarity may also suggest a control of membrane permeability, hence increasing the porosity during planktonic growth however, with the prediction of cytoplasmic localization (97%) using SubLoc, this may be a transcription control for CadC. Further analysis is warranted to assess this prediction.

These predictions will be used as a basis for analyzing phenotypical and morphological changes post-mutagenesis. At this time it is suggested with removal of b1995, biofilm cells will develop reduced resistance to shear forces and toxic substances due to unregulated outer membrane porosity. Other affects after removal the removal of b1995 are undeterminable due to the proposed trans-acting nature of the gene product

hence the further genomic affects. The removal of b1394 may reduce the utilization of certain carbon sources under anaerobic growth, hence reduced viability of both biofilm and planktonic phenotypes. Further analysis is needed to assess the post-mutagenesis changes for b2852, however at this point, predicted changes are in transcript regulation of the outer-membrane control gene products. The removal of these genes of interest should produce clear characterization results, as the potential for removal of other open reading frames and hence further genomic affects, should not be a concern. However, one should be aware of the potential genomic affects due to removal of the trans-acting b1995.

This study may lead to the control of biofilm formation and viability in industry, medicine, and the environment. One may use these genes or gene products as potential targets for toxic substances. As with b1995, one could target transcriptional or translation control and hence reduce the resistance to antibiotics or other toxins based on membrane porosity. Post-mutagenesis analyses should include biofilm formation in once through flow cell systems and further treatment with antibiotics and varied nutrients. From this study, one can conclude that pre-mutagenesis sequence analyses truly aid in the prediction of complications associated with knockout mutations and determination of the reliability of mutagenesis results.

LITERATURE CITED

BUCHET, A., K. EICHLER and M.-A. MANDRAND-BERTHELOT, 1998 Regulation of the Carnitine Pathway in Escherichia coli: Investigation of the cai-fix Divergent Promoter Region. J. Bacteriol. **180:** 2599-2608.

DE WULF, P., A. M. MCGUIRE, X. LIU and E. C. C. LIN, 2002 Genome-wide Profiling of Promoter Recognition by the Two-component Response Regulator CpxR-P in Escherichia coli. J. Biol. Chem. **277:** 26652-26661.

DIAZ, E., A. FERRANDEZ, M. A. PRIETO and J. L. GARCIA, 2001 Biodegradation of Aromatic Compounds by Escherichia coli. Microbiol. Mol. Biol. Rev. **65:** 523-569.

FERRANDEZ, A., B. MINAMBRES, B. GARCIA, E. R. OLIVERA, J. M. LUENGO *et al.*, 1998 Catabolism of Phenylacetic Acid in Escherichia coli. CHARACTERIZATION OF A NEW AEROBIC HYBRID PATHWAY. J. Biol. Chem. **273:** 25974-25986.

FUSON, K. L., 2002 Biofilm-induced gene expression in chemostat grown *Escherichia coli* as determined by a gene array, pp. 60 in *Biology*. Texas State University, San Marcos, Texas.

MOLLER, S., C. STERNBERG, J. B. ANDERSEN, B. B. CHRISTENSEN, J. L. RAMOS *et al.*, 1998 In Situ Gene Expression in Mixed-Culture Biofilms: Evidence of Metabolic Interactions between Community Members. Appl. Environ. Microbiol. **64:** 721-732.

NIKAIDO, H., 1996 Outer membrane, pp. 29-47 in *Escherichia coli and Salmonella: cellular and molecular biology*, edited by H. E. UMBARGER. ASM Press, Washington, D.C.

NUSSBAUM-SHOCHAT, A., and O. AMSTER-CHODER, 1999 BglG, the transcriptional antiterminator of the bgl system, interacts with the beta ' subunit of the Escherichia coli RNA polymerase. PNAS **96:** 4336-4341.

SAMARTZIDOU, H., and A. H. DELCOUR, 1999 Excretion of Endogenous Cadaverine Leads to a Decrease in Porin-Mediated Outer Membrane Permeability. J. Bacteriol. **181:** 791-798.

CHAPTER III

CORRELATION BETWEEN COLANIC ACID PRODUCTION AND BIOFILM FORMATION

INTRODUCTION

Biofilms, surface-adherent microorganisms, are important factors in many of today's environments. They are culprits of prosthetic and indwelling medical device infections, causative agents of pipeline contamination, and beneficial characters in bioremediation and water ecology stability. Today, much research focuses on the underlying question of which cellular and genomic factors influence biofilm formation, viability, and dispersion (DAVEY and O'TOOLE 2000; O'TOOLE *et al.* 1999; STOODLEY *et al.* 2002). Those cellular and genomic factors influence a variety of biochemical compounds.

Colanic acid has been shown as a major component of *Escherichia coli* biofilms (DANESE *et al.* 2000). The biochemical properties of colanic acid have been described in depth, while the role of colanic acid in biofilm formation is still unknown. Colanic acid was first described by Walther Goebel as a serologically active, nitrogen-free polysaccharide (Geobel 1963). Goebel was also the first to describe colanic acid as an M antigen (Geobel 1963). Colanic acid, the M-antigen, differs from the more common

but closely related O-antigen, or lipopolysaccharide, by its high molecular weight and production of a capsule rather than slime lipopolysaccharide (LPS), as is the case with the latter. Colanic acid has been shown repeatedly to contain the sugars glucose, galactose, fucose, and glucuronic acid (ANDERSON and ROGERS 1963; GRANT *et al.* 1969). When colanic acid is produced at increased concentrations the cells appear mucoid and colonies are enlarged in diameter as compared to colonies undergoing normal production (Geobel 1963). Colanic acid itself has also been proposed as a protective agent for bacteria when they are outside the host intestines or under desiccant stress (Ophir and Gutnick 1994).

Hence, understanding control of colanic acid may lead to a better comprehension of biofilms themselves. The production of colanic acid occurs by formation of undecaprenol pyrophosphate-linked oligosaccharide repeat units at the cytoplasmic face of the inner membrane. These chains are then transported across the inner membrane by a Wzx protein-dependent process followed by polymerization involving a Wzy protein (Drummelsmith and Whitfield 2000). This exopolysaccharide has been shown to be produced in a regulatory fashion so that mucoid colonies are not constitutively expressed (GRANT *et al.* 1969).

Colanic acid itself is produced via regulation of a gene cluster known as the *wca* region (or *cps* region) (STEVENSON *et al.* 1996). This cluster contains 19 different genes, totaling 23 kb, that are hypothetically transcribed as one unit and followed by an operon encoding enzymes that produce the O-antigen (STEVENSON *et al.* 1996). See Figure 2 for an illustration of the *wca* operon. The *wca* genes produce products involved in assembly, transport, and modification of colanic acid. Three regulatory genes control transcription of the *wca* genes, *rcsA*, *rcsB*, and *rscC* (GOTTESMAN *et al.* 1985). The first two of these

three genes act positively, but unstably,, to increase transcription of the *wca* genes and are further regulated by MdoH (EBEL and TREMPY 1999; EBEL *et al.* 1997). These proteins act unstably due to the ability to be degraded by *lon,* an indirect negative transcriptional regulator for *wca* (GOTTESMAN *et al.* 1985). The latter of the three genes, *rcsC,* acts as an unstable negative regulator of *wca* transcription (GOTTESMAN *et al.* 1985). An illustration of the biosynthesis pathway for colanic acid is shown in Figure 3, and was derived from Stevenson *et al.* (STEVENSON *et al.* 1996).

Wzc is an inner membrane protein with an ATP-binding domain and three hypothesized transmembrane segments (STEVENSON *et al.* 1996). Wzc has been found essential for the assembly of the high molecular weight group I capsular polysaccharide, colanic acid, on the surface of *E. coli* (Drummelsmith and Whitfield 1999). Absence of Wzc results in impaired surface expression of colanic acid, however, oligosaccharides with several repeat units are still synthesized and assembled on the surface as lipopolysaccharide, constituting the lipid-A core (Drummelsmith and Whitfield 2000). Wzc act as a protein-tyrosine kinase which is able to autophosphorylate when overproduced and has a specific dephosphorylation cognate, Wzb (VINCENT *et al.* 1999). Wzc has remarkable sequence similarity to Ptk, a protein-tyrosine kinase of *Acinetobacter johnsonii,* which also has autophosphorylation capabilities on tyrosine and a cognate phosphatase, Ptp(VINCENT *et al.* 1999). From these data it is suggested that *wzc*⁻ mutants would lack the M-antigen exopolysaccharide, colanic acid, express LPS, and consequently form biofilms differently than their parental counterparts.

As with much of the work characterizing components involved with biofilms, the situation is complex. The role exopolysaccharides, namely colanic acid, play in biofilm formation is unknown. Prigent-Combaret *et al.* have shown that colanic acid production was not induced until the cells had reached the attachment stage (PRIGENT-COMBARET *et al.* 1999). While other studies have shown that exopolysaccharide production by Gram-negative organisms is necessary for surface attachment of the biofilm (Watnick and Kolter 1999). This study aims to elucidate the role of colanic acid deficiencies, production, and overproduction given parentals and mutants of the *wzc* gene. To address those questions, biofilm production, colanic acid expression, and growth rates were compared to assess the correlation of colanic acid production and biofilms.

MATERIALS AND METHODS

Strains and media. Escherichia coli JM83 and JM83wzc::Km (VINCENT *et al.* 2000; YANISCH-PERRON *et al.* 1985) were obtained from Christophe Grangeasse, Université de Lyon. Plasmids, pUC18wzc-rcsA$^+$, pUC18wzc$_{569}$-rcsA$^+$, pUC18wzc$_{L6}$-rcsA$^+$, pUC18wzc$_{569L6}$-rcsA$^+$, pUC18wzc$_{K540}$-rcsA$^+$, and pUC18-rcsA$^+$ (DOUBLET *et al.* 2002; VINCENT *et al.* 2000) were also obtained from Christophe Grangeasse. All strains were grown in Luria-Bertani (rich medium) or M63 with 0.1% w/v casamino acids, 20ug/mL thiamine, and 0.2% glucose as a carbon source (minimal medium) (MILLER 1972). Strains were grown at 37°C with aeration unless otherwise specified. Kanamycin was utilized at 25ug for all JM83wzc::Km strains and ampicillin at 100ug for retention of plasmids in transformed strains. High resistance water, obtained from a Milli-Q Plus™

water purification system (Millipore, Bedford, MA), was used throughout this study unless otherwise specified.

Transformation of wild type and mutant JM83. *E. coli* JM83 and JM83wzc::Km were transformed with each plasmid described above. Cells were grown overnight in Luria-Bertani broth then a 1% v/v subculture was grown to an OD_{600} between 0.5 and 0.6. Cells were kept on ice from this point forward. Cells were centrifuged at 6,000g, 4°C, for six minutes. Cells were washed four times with 10% ice-cold glycerol then resuspended in $1/125^{th}$ the volume of original culture. Cells were stored at –80°C until use but no longer than six months. See Table 1 for new strain designations. Plasmid presence was verified by extraction with QIAprep Spin Miniprep Kit (Qiagen, Maryland, USA, cat. no. 27106) followed by 1.0% Seakem LE agarose gel electrophoresis. Bands were visualized after the gel was stained for 30 minutes in 0.5ug/mL ethidium bromide and destained for 30 minutes in double deionized water.

Verification of colanic acid production. Colanic acid was extracted via previously described procedures (CERNING *et al.* 1994). For clarity, procedures are as follows. Twice subcultured cells were inoculated in 1% v/v concentration to 50mL of nutrient rich and minimal media, grown for 48 hours at 37°C with aeration then heated for 15 minutes at 100°C to denature EPS-degrading enzymes. Cells were cooled to 4°C and centrifuged at 4°C at 13,200g for 30 minutes. Forty milliliters of supernatant was precipitated by addition of three volumes of 100% reagent grade ethanol and stored at 4°C overnight. Mixture was centrifuged as described above. The pellet was dissolved in

5mL of Millipore water, dialyzed for 48 hours against Millipore water in Spectra/Por (Spectrum, California, USA, catalog 132720) dialysis tubing with an MWCO of 3,500 Da and then lyophilized. Residual polypeptides were removed by precipitation with 5mL of 10% v/v trichloroacetic acid and centrifuged as before. Supernatant was re-dialysed for 5 days against Millipore water and dried. Resulting compound was re-suspended in 1 mL Millipore water and stored at 4°C until quantification. Quantification of colanic acid was carried per previously described procedures (Dische and Shettles 1951). Once again for clarity, 1 mL of the colanic acid extract was mixed with 4.5 mL of H_2SO_4/H_2O (6:1, by volume). Mixture was warmed to room temperature then heated at 100°C for 20 minutes and once again cooled to room temperature. Absorbances were measured at 396nm and 427nm either directly or after addition of 900uL of 3% cysteine hydrochloride to 50uL of sample. Resulting absorbances were directly correlated to non-dialyzable methylpentose concentration and compared to fucose ranging from 5 to 95mg/mL as standards.

Quantification of EPS production by Biofilms. Residual attached exopolysaccharide (EPS) was utilized for biofilm quantification as described previously (O'TOOLE *et al.* 1999). All strains were subcultured twice in respective media (nutrient rich and minimal) prior to assay. From a fresh overnight broth culture (growth without shaking), 50uL of suspension was inoculated into 5mL of fresh media then 200uL were transferred to 10 replicate wells in a 96-well polystyrene culture plate. Negative controls were maintained with respective fresh media and prepared in the same manner as cultures. Plates were covered and incubated at 37°C for 48 hours in a plastic box with small amount of Millipore water for humidity. Growth was assayed by absorbance at

570nm on a Bio-Tek EL311 Autoreader. From a 0.1% freshly prepared crystal violet mixture, 10uL were added to each well and allowed to sit for 15 minutes at room temperature. Excess stain and unattached cells were removed then wells were washed twice with sterile Millipore water. Remaining crystal violet was then solubilized in 200uL of 100% ethanol. Absorption of crystal violet was recorded at 570nm. Analysis of covariance was performed to assess differences between EPS production, with culture density as a covariate, in differing media and strains. A two-factor Analysis of Covariance (ANCOVA) with type three sums of squares (due to missing culture density values) was performed with media and strain as factors, culture optical density as the covariate, and crystal violet (EPS) absorbance as the response variable. Biolfilm optical densities were transformed by addition of one to all values. This transformation corrected failures to meet the assumptions for homoscedasticity required for the analysis of covariance. Multiple regression analyses were performed to assess differences in EPS absorbance means corrected for the covariate.

RESULTS

Transformation. All strains were transformed efficiently with positive growth on selective media. See Table 1 for antibiotic selection requirements along with strain descriptions. Negative control transformations (parental strains only with no plasmids) did not have positive growth on selective media. Plasmid preparations yielded positive presence of plasmid in all strains transformed. Negative transformation reactions did not yield visible presence of plasmid when visualized on a 1.0% Seakem LE agarose gel after staining with ethidium bromide.

Verification of colanic acid production. Statistical analyses were not performed on colanic acid extracts because replicates were not performed. However, extract concentrations were compared to standard concentrations of fucose. In general, colanic acid production was greater in LB cultures than M63 with 0.2% glucose cultures and may be attributed to increased growth. Only diluted extracts were considered due to maximum optical density limits set forth by equipment. Highest amounts of biological extract (presumably fucose from colanic acid) were produced by CLB114, CLB115, CLB116, and CLB118. This was apparent in both media types. CLB111 and CLB112 produced the least amount of biological extract in M63 with 0.2% glucose cultures where CLB108, CLB117, and CLB119 produced the least amount in LB. These differences did not seem attributed to culture densities within media type. CLB108, CLB109, CLB110, CLB112, CLB113, CLB117, CLB118, CLB119, JM83, and JM83wzc::KmR all produced less than 5mg of fucose. CLB111 produced close to 5mg. CLB114, CLB115, and CLB116 produced greater than 5mg of fucose, but less than 15mg.

Quantification of EPS production by biofilms. Differences in growth amounts were seen between media and strain with a significant p value of zero. Differences between culture densities were insignificant ($p=0.13$), however, interaction between strain and culture density was significant. These differences were further analyzed by a multiple regression analysis. Further significant interaction was noted with media and strain. Analysis of covariance suggested indicated differences between absorption values of crystal violet due to EPS production with respect to media and also strain ($p=0.00$ and $p=0.00$, respectively). Summary of the ANCOVA is presented in Table 4. Multiple regression analysis indicated that significant interaction between strain and culture

density may be a result of relatively high biofilm production by CLB110, CLB112, and CLB118. Multiple regression accounted for little more than half of the variation in EPS between strains (r^2=0.5981) indicating that culture density was not the only variable leading to differences between strains. When culture density was taken into account as a covariate, CLB110 produced the most EPS followed by CLB112, then CLB118 and CLB109. Further trend is illustrated in Table 2. A common trend was observed with respect to biofilm formation between both media. Mean values for culture and biofilm optical density along with standard deviations are given in Table 2.

DISCUSSION

Contradicting studies have stated that colanic acid is not induced until late maturity, prevents adherence, and aids in adherence with respect to biofilm formation (PRIGENT-COMBARET *et al.* 1999; WATNICK and KOLTER 1999). The data presented in this study aim to address these contradictory findings.

Colanic acid quantification revealed no correlation between growth densities and amount of non-dialyzable methyl-pentose, namely fucose. Increases in colanic acid production when cells were cultured in rich media may be attributed to increased energy (ATP) production by the cell. ATP presence is required for phosphorylation reactions leading to expression of surface colanic acid (STEVENSON *et al.* 1996). Increased biological extracts due to media composition are negated by the control extraction where absorbances at 427nm where identical in nutrient rich and minimal media. Amounts of colanic acid produced did not correlate to the functionality of *wzc* suggesting a lack of influence on EPS production after 48 hours of growth, a relatively young biofilm as

described in Chapter 1. Correlation did occur between growth characteristics on solid media and the amount of colanic acid. This suggests that measured amounts of colanic acid were quantifiable with respect to strains but it must be noted that carry-over loss does indeed occur during extraction processes.

This study found that biofilms followed the same trend with respect to EPS production when grown in nutrient rich and minimal media. Biofilm production was reduced when cells were grown in minimal media, which may be attributed to stringent response by the cell when under stress (Balzer and McLean 2002) or could also be attributed to ATP production as was proposed for colanic acid production. This significance is demonstrated by the ANCOVA. Significant differences in EPS production were also found between strains, suggesting that the mutations or additions of plasmids do have an affect on EPS production. Given that CLB108 and JM83wzc::KmR (parental of CLB108) produce almost identical amounts of EPS, it is suggested that the mutation in *wcz* is not the contributing factor for differences in EPS production. From this it is further suggested that decreases in phosphorylation and catalytic site activity may promote EPS production. Multiple regression analysis illustrated differences in EPS production while accounting for variance in culture densities enabling one to negate the factor of growth restrictions or increases due to transcription speed of plasmid and chromosome along with affects of different antibiotic selections.

The methodologies herein did not assess biofilm formation under sheer environments or after full community maturity, which are commonly attributed to the formation of adherent populations. Therefore, it is possible that expression of the *wca* operon and its regulator, *rcsA* did not directly attribute to biofilm formation but rather to

only EPS production. There was no clear correlation between biofilm (EPS) production and colanic acid production except for where phosphorylation and catalytic site activity were reduced (mutated) colanic acid and EPS productions were increased. From this, it is suggested that *wzc,* of the colanic acid operon, does not have a significant influence on adherent biofilm (EPS) formation during the first 48 hours of growth. Given that measurements of EPS production were only taken for that which remained attached to the polystyrene one cannot truly assess if EPS was purely colanic acid or some other type of polysaccharide that may include LPS. A previous studied showed, through atomic force microscopy, that wild-type and over-producers for colanic acid do not adhere to surfaces (HANNA *et al.* 2003). Those authors suggested that colanic acid may interfere or block time-dependent adhesion of bacteria to hydrophobic and hydrophilic surfaces (HANNA *et al.* 2003). Time dependent adherence may have led to lack of correlation between biofilm and colanic acid production given measurements only accounted for residual EPS produced after 48 hours and not after full biofilm maturity, which occurs after six days of growth (shown in Chapter 1). Another study demonstrated that EPS production, namely colanic acid, protected cells while under desiccant stress (Ophir and Gutnick 1994). This was supported by the current investigation where wells that harbored strains with increased EPS production did not desiccate after one week at room temperature while others were completely desiccated. This was noted after the stained EPS was solubilized in 100% ethanol (data not shown).

Further study is warranted to assess the role of these mutations over time, namely using chemostat and flow cell reactors. An increase in volume during biofilm production

and colanic acid measurements may result in less variance and higher yields of EPS and colanic acid hence, detection limits become negligible. Negating the detection limits may also lead to increased correlation between biofilm EPS and colanic acid production.

LITERATURE CITED

ANDERSON, E. S., and A. H. ROGERS, 1963 Slime polysaccharides of the *Enterobacteriaceae*. Nature **198:** 714-715.

BALZER, G. J., and R. J. C. MCLEAN, 2002 The stringent response genes *relA* and *spoT* are important for *Escherichia coli* biofilms under slow-growth conditions. Canadian Journal of Microbiology **48:** 675-680.

CERNING, J., C. RENARD, J. THIBAULT, C. BOUILLANNE, M. LANDON *et al.*, 1994 Carbon Source Requirements for Exopolysaccharide Production by Lactobacillus casei CG11 and Partial Structure Analysis of the Polymer. Appl. Environ. Microbiol. **60:** 3914-3919.

DANESE, P. N., L. A. PRATT and R. KOLTER, 2000 Exopolysaccharide Production Is Required for Development of Escherichia coli K-12 Biofilm Architecture. J. Bacteriol. **182:** 3593-3596.

DAVEY, M. E., and G. A. O'TOOLE, 2000 Microbial biofilms: from ecology to molecular genetics. Microbiology and Molecular Biology Reviews **64:** 847-867.

DISCHE, Z., and L. B. SHETTLES, 1951 A New Spectrophotometric Test for the Detection of Methylpentose. J. Biol. Chem. **192:** 579-582.

DOUBLET, P., C. GRANGEASSE, B. OBADIA, E. VAGANAY and A. J. COZZONE, 2002 Structural Organization of the Protein-tyrosine Autokinase Wzc within Escherichia coli Cells. J. Biol. Chem. **277:** 37339-37348.

DRUMMELSMITH, J., and C. WHITFIELD, 1999 Gene products required for surface expression of the capsular form of the group 1 K antigen in Escherichia coli (O9a:K30). Mol Microbiol **31:** 1321-1332.

DRUMMELSMITH, J., and C. WHITFIELD, 2000 Translocation of group 1 capsular polysaccharide to the surface of Escherichia coli requires a multimeric complex in the outer membrane. EMBO J. **19:** 57-66.

EBEL, W., and J. E. TREMPY, 1999 Escherichia coli RcsA, a Positive Activator of Colanic Acid Capsular Polysaccharide Synthesis, Functions To Activate Its Own Expression. J. Bacteriol. **181:** 577-584.

EBEL, W., G. VAUGHN, H. PETERS, 3RD and J. TREMPY, 1997 Inactivation of mdoH leads to increased expression of colanic acid capsular polysaccharide in Escherichia coli. J. Bacteriol. **179:** 6858-6861.

GEOBEL, W., 1963 Colanic acid. PNAS **49:** 464-471.

GOTTESMAN, S., P. TRISLER and A. TORRES-CABASSA, 1985 Regulation of capsular polysaccharide synthesis in *Escherichia coli* K-12: characterization of three regulatory genes. J. Bacteriol. **162:** 1111-1119.

GRANT, W. D., I. W. SUTHERLAND and J. F. WILKINSON, 1969 Exopolysaccharide colanic acid and its occurrence in the *Enterobacteriaceae*. J. Bacteriol. **100:** 1187-1193.

HANNA, A., M. BERG, V. STOUT and A. RAZATOS, 2003 Role of Capsular Colanic Acid in Adhesion of Uropathogenic Escherichia coli. Appl. Environ. Microbiol. **69:** 4474-4481.

MILLER, J., 1972 *Experiments in molecular genetics*. Cold Spring Harbor Laboratory PRess, Cold Spring Harbor, NY.

OPHIR, T., and D. L. GUTNICK, 1994 A role for exopolysaccharide in the protection of microorganisms from desiccation. Appl. Environ. Microbiol. **60:** 740-745.

O'TOOLE, G. A., L. A. PRATT, P. I. WATNICK, D. K. NEWMAN, V. B. WEAVER *et al.*, 1999 Genetic approaches to study of biofilms. Methods Enzymol **310:** 91-109.

PRIGENT-COMBARET, C., O. VIDAL, C. DOREL and P. LEJEUNE, 1999 Abiotic surface sensing and biofilm-dependent regulation of gene expression in *Escherichia coli.* J. Bacteriol. **181:** 5993-6002.

STEVENSON, G., K. ANDRIANOPOULOS, M. HOBBS and P. REEVES, 1996 Organization of the Escherichia coli K-12 gene cluster responsible for production of the extracellular polysaccharide colanic acid. J. Bacteriol. **178:** 4885-4893.

STOODLEY, P., K. SAUER, D. G. DAVIES and J. W. COSTERTON, 2002 Biofilms as complex differentiated communities. Annual Review of Microbiology **56:** 187-209.

VINCENT, C., P. DOUBLET, C. GRANGEASSE, E. VAGANAY, A. J. COZZONE *et al.*, 1999 Cells of Escherichia coli Contain a Protein-Tyrosine Kinase, Wzc, and a Phosphotyrosine-Protein Phosphatase, Wzb. J. Bacteriol. **181:** 3472-3477.

VINCENT, C., B. DUCLOS, C. GRANGEASSE, E. VAGANAY, M. RIBERTY *et al.*, 2000 Relationship between exopolysaccharide production and protein-tyrosine phosphorylation in gram-negative bacteria. Journal of Molecular Biology **304:** 311-321.

WATNICK, P. I., and R. KOLTER, 1999 Steps in the development of *Vibrio cholerae* El Tor biofilm. Mol. Microbiol. **34:** 586-595.

YANISCH-PERRON, C., J. VIEIRA and J. MESSING, 1985 Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mpl8 and pUC19 vectors. Gene **33:** 103-119.

APPENDIX I – CHAPTER I TABLES AND FIGURES

Table 1. Three-factor analysis of variance results for *Escherichia coli* K-12 gene expression of four and six day chemostat grown biofilm cultures. Maturity and gene function are independent while gene was nested in gene function and random. Data were log transformed to meet assumptions.

Factor	Df	Mean Square	F value	P value
Maturity	2	6.044274	33.51203	<0.0001
Gene Function	17	0.311247	1.72569	0.0387901
Gene in Gene Function	108	0.307560	1.70524	0.003397
Maturity:Gene Function	17	0.261870	1.45192	0.1131431
Maturity:Gene within Gene Function	108	0.269568	1.49460	0.0054328
Residuals	251	0.180361		

Table 2. Results of Tukey's HSD for *Escherichia coli* K-12 gene expression responses in all gene function categories between four and six day biofilm maturity levels. Eighteen Tukey's were performed, one for each function category to assess expression differences between maturity.

Gene Function	4 Day Mean	6 Day Mean	Significance
Translation, Ribosomal Structure and Biogenesis	1.02522	1.53919	Yes
Transcription	1.10107	1.48359	Yes
DNA Replication, Recombination and Repair	1.19662	1.96419	Yes
Cell Division and Chromosome Partitioning	1.24213	3.19495	No
Posttranslational Modification, Protein Turnover, and Chaperones	1.26375	1.26935	No
Cell Envelope Biogenesis and Outer Membrane	1.16107	1.53847	Yes
Cell Motility and Secretion	0.91362	1.41427	No
Inorganic Ion Transport and Metabolism	1.27188	1.92332	No
Signal Transduction Mechanisms	1.12021	1.18192	No
Energy Production and Conversion	1.05078	1.59171	Yes
Carbohydrate Transport and Metabolism	1.21441	2.60285	Yes
Amino Acid Transport and Metabolism	1.10718	2.10374	No
Nucleotide Transport and Metabolism	1.25947	1.51208	No
Coenzyme Metabolism	1.00416	1.71948	Yes
Lipid Metabolism	1.09364	1.83013	Yes
Secondary Metabolites Biosynthesis and Transport Catabolism	1.11485	1.49834	Yes
General Function-Prediction Only	1.36012	1.88722	No
Function Unknown	1.18806	1.11511	No

Table 3. Descriptions for genes used as gene factor within gene function category for ANOVA. See Figure 1 for a description of gene function categories. Descriptions and accessions were obtain from NCBI COGs Functional Annotation web database found at http://www.ncbi.nlm.nih.gov/COG/old/palox.cgi?fun=all. Those without descriptive information have changed gene names or have been re-categorized since research was performed.

Gene Function Category	Gene	Chromosomal Position	Actual Function	NCBI Accession
A	glcG	3121843-3122247	Uncharacterized protein, possibly involved in utilization of glycolate and propanediol	COG3193
A	tehB	1499586-1500179	Uncharacterized protein, possibly involved in tellurite resistance	COG3615
A	dedA	2432102-2432761	Uncharacterized membrane-associated protein	COG0586
A	relE	1643657-1643370	Uncharacterized ACR	GOG2026
A	bacA	3201772-3200951	Uncharacterized ACR, bacitracin resistance protein	COG1968
A	erfK	2061345-2060413	31.6 kD protein in cobT 3'region precursor	
A	sprT	3820027-3822135	Uncharacterized BCR	COG3091
B	rhaT	4098106-4097072	Permeases of the drug/metabolite transporter (DMT) superfamily	COG0697
B	thiH	4189446-4188313	Thiamine biosynthesis enzyme ThiH and related uncharacterized enzymes	COG1060
B	rimI	1124785-1125369	Acetyltransferases	COG0456
B	tldD	3389663-3388218	Predicted Zn-dependent proteases and their inactivated homologs	COG0312
B	comFC		Predicted amidophosphoribosyltransferases	COG1040
B	rcsA	2021990-2022613	positive regulator for ctr capsule biosynthesis, positive transcription factor	
B	dcuB	4346322-4344982	Anaerobic C4-dicarboxylate transporter	COG2704
C	menF	2378438-2377368	Isochorismate synthase	COG1169
C	fabG	1149893-1150627	Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)	COG1028
C	mdlB	469860-471641	ABC-type multidrug/protein/lipid transport system, ATPase component	COG1132
C	entB	626917-627774	Isochorismate hydrolase	COG1535
C	sufI	3159273-3160685	Putative multicopper oxidases	COG2132
C	emrA	2809448-2810620	Multidrug resistance efflux pump	COG1566
C	smtA	972760-973545	SAM-dependent methyltransferases	COG0500
D	atoD	2321467-2322129	Acyl CoA:acetate/3-ketoacid CoA transferase, alpha subunit	COG1788
D	caiA	39244-40386	Acyl-CoA dehydrogenases	COG1960

D	fadL	2459320-2460666	Long-chain fatty acid transport protein	COG2067
D	pspA	1366103-1366771	Phage shock protein A (IM30), suppresses sigma54-dependent transcription	COG1842
D	sbm	3058870-3061014	Methylmalonyl-CoA mutase, N-terminal domain/subunit	COG1884
D	fabH	1147982-1148935	3-oxoacyl-[acyl-carrier-protein] synthase III	COG0332
D	acpS	2698638-2699018	Phosphopantetheinyl transferase (holo-ACP synthase)	COG0736
E	menA	4117003-4117929	1,4-dihydroxy-2-naphthoate octaprenyltransferase	COG1575
E	ubiX	2426077-2426646	3-polyprenyl-4-hydroxybenzoate decarboxylase	COG0163
E	bioF	809604-810758	7-keto-8-aminopelargonate synthetase and related enzymes	COG1056
E	birA	4170661-4171626	Biotin-(acetyl-CoA carboxylase) ligase	COG0340
E	hemB	387977-388984	Delta-aminolevulinic acid dehydratase	COG0113
E	folA	49823-50302	Dihydrofolate reductase	COG0262
E	btuC	1792196-1793176	ABC-type cobalamin/Fe3+-siderophores transport systems, permease components	COG0609
F	add	1700257-1701258	Adenosine deaminase	COG1816
F	purN	2620254-2620892	Folate-dependent phosphoribosylglycinamide formyltransferase PurN	COG0299
F	gmk	3819055-3819678	Guanylate kinase	COG0194
F	nupC	2511062-2512264	Nucleoside permease	COG1972
F	pyrE	3812754-3813395	Orotate phosphoribosyltransferase	COG0461
F	prsA	1260151-1261098	Phosphoribosylpyrophosphate synthetase	COG0462
F	codB	354146-355405	Purine-cytosine permease and related proteins	COG1457
G	aroA	958035-959318	3-Deoxy-D-arabino-heptulosonate 7-phosphate (DAHP) synthase	COG2876
G	livG	3591070-3591837	ABC-type branched-chain amino acid transport systems, ATPase component	COG0411
G	potB	1182840-1183667	ABC-type spermidine/putrescine transport system, permease component I	COG1176
G	pepB		Leucyl aminopeptidase	COG0260
G	carB	30817-34038	Carbamoylphosphate synthase large subunit (split gene in MJ)	COG0458
G	glyA	2682274-2683527	Glycine hydroxymethyltransferase	COG0112
G	leuA	81958-83529	Isopropylmalate/homocitrate/citramalate synthases	COG0119
H	malG	4240205-4241095	ABC-type maltose transport systems, permease component	COG3833
H	lacZ	362455-365529	Beta-galactosidase/beta-glucuronidase	COG3250

H	uxaC	3240969-3242381	Glucuronate isomerase	COG1904
H	celC	1817479-1817829	Phosphotransferase system cellobiose-specific component IIA	COG1447
H	otsB	1979611-1980411	Trehalose-6-phosphatase	COG1877
H	xylB	3725546-3727000	Sugar (pentulose and hexulose) kinases	COG1070
H	kduI	2981310-2982146	5-keto 4-deoxyuronate isomerase	COG3717
I	bisC	3711690-3713909	Anaerobic dehydrogenases, typically selenocysteine-containing	COG0243
I	fixB	43188-44129	Electron transfer flavoprotein alpha-subunit	COG2025
I	atpC	3913181-3913600	F0F1-type ATP synthase epsilon subunit (mitochondrial delta subunit)	COG0355
I	fldA	710158-710688	Flavodoxins	COG0716
I	cyoC	447270-447884	Heme/copper-type cytochrome/quinol oxidase, subunit 3	COG1845
I	ndh	1165308-1166612	NADH dehydrogenase, FAD-containing subunit	COG1252
I	eutG	2566344-2567558	Alcohol dehydrogenase IV	COG1454
J	spoT	3820027-3822135	Guanosine polyphosphate pyrophosphohydrolases/synthetases	COG0317
J	fecR	4514332-4515285	Fe2+-dicitrate sensor, membrane component	COG3712
J	bolA	453663-454013	Stress-induced morphogen (activity unknown)	COG0271
J	atoS	2318063-2319889	PAS/PAC domain	COG2202
J	marQ			
J	uspA	3637741-3638175	Universal stress protein UspA and related nucleotide-binding proteins	COG0589
J	kdpD	720953-723637	Osmosensitive K+ channel histidine kinase	COG2205
K	btuC	1792196-1793176	ABC-type cobalamin/Fe3+-siderophores transport systems, permease components	COG0609
K	modA	794312-795085	ABC-type molybdate transport system, periplasmic component	COG0725
K	pstA	3906177-3907067	ABC-type phosphate transport system, permease component	COG0581
K	cirA	2242798-2244789	Outer membrane receptor proteins, mostly Fe transport	COG1629
K	sodA	4098391-4099011	Superoxide dismutase	COG0605
K	amtB	472190-473476	Ammonia permeases	COG0004
K	trkA	3434155-3435531	K+ transport systems, NAD-binding component	COG0569
L	malK	4244363-4245478	ABC-type sugar transport systems, ATPase components	COG3839
L	cheW	1970860-1971363	Chemotaxis signal transduction protein	COG0835

L	fliP	2019891-2020628	Flagellar biosynthesis/type III secretory pathway protein	COG1338
L	lspA	25207-25701	Lipoprotein signal peptidase	COG0597
L	fimD	4542665-4545301	PapC-like porin protein involved in fimbrial biogenesis	COG3188
L	ftsY	3600381-3601874	Signal recognition particle GTPase	COG0552
L	flgM	1129058-1129351	Negative regulator of flagellin synthesis (anti-sigma-28 factor)	COG2747
M	lpxA	202560-203348	Acyl-[acyl carrier protein]--UDP-N-acetylglucosamine O-acyltransferase	COG1043
M	murI	4162995-4163864	Glutamate racemase	COG0796
M	mdoH	1110086-1112629	Membrane glycosyltransferases	COG2943
M	tolC	3176125-3177612	Outer membrane protein	COG1538
M	tonB	1309113-1309832	Periplasmic protein TonB, links inner and outer membranes	COG0810
M	slp	3651558-3652157	Starvation-inducible outer membrane lipoprotein	COG3065
M	murD	97087-98403	UDP-N-acetylmuramoylalanine-D-glutamate ligase	COG0771
N	clpX	456650-457924	ATP-dependent protease Clp, ATPase subunit	COG1219
N	ccmE	2292921-2293400	Cytochrome c-type biogenesis protein CcmE	COG2332
N	cbpA	1062078-1062998	Molecular chaperones, DnaJ class	COG2214
N	ahpC	638168-638731	Peroxiredoxin	COG0450
N	pflA	949563-950303	Pyruvate-formate lyase-activating enzyme	COG1180
N	sppA	1846861-1848717	Periplasmic serine proteases (ClpP class)	COG0616
N	nrdG	4457474-4457938	Organic radical activating enzymes	COG0602
O	mesJ	212331-213629	Predicted ATPase of the PP-loop superfamily implicated in cell cycle control	COG0037
O	ftsE	3599710-3600378	Predicted ATPase involved in cell division	COG2884
O	minC	1224608-1225303	Septum formation inhibitor	COG0850
O	mukF	973542-974864	Uncharacterized protein involved in chromosome partitioning	COG3006
O	ftsZ	105305-106456	Cell division GTPase	COG0206
O	mrp	2191079-2192218	ATPases involved in chromosome partitioning	COG0489
O	mreB	3397681-3398784	HSP70 class molecular chaperones involved in cell morphogenesis	COG1077
P	hepA	60358-63264	Superfamily II DNA/RNA helicases, SNF2 family	COG0553
P	recD	2948657-2950483	ATP-dependent exoDNAse (exonuclease V), alpha subunit - helicase superfamily I member	COG0507

P	holD	4605372-4605785	DNA polymerase III, psi subunit	COG3050
P	phrB	738730-740148	Deoxyribodipyrimidine photolyase	COG0415
P	seqA	712210-712755	Negative regulator of replication initiationR	COG3057
P	uvrB	812749-814770	Helicase subunit of the DNA excision repair complex	COG0556
P	topA	1329072-1331669	Topoisomerase IA	COG0550
Q	fliA	1999093-1999812	DNA-directed RNA polymerase specialized sigma subunit	COG1191
Q	araC	70387-71265	AraC-type DNA-binding domain-containing proteins	COG2207
Q	cadC	4357974-4359512	DNA-binding winged-HTH domains	COG3710
Q	rpoZ	3819733-3820008	DNA-directed RNA polymerase subunit K/omega	COG1758
Q	vacB	4404190-4406673	Exoribonucleases	COG0557
Q	pspC	1367049-1367408	Putative stress-responsive transcriptional regulator	COG1983
Q	ompR	3533503-3534222	Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain	COG0745
R	rimL	1496962-1497501	Acetyltransferases, including N-acetylases of ribosomal proteins	COG1670
R	gatA	2172617-2173069	Asp-tRNAAsn/Glu-tRNAGln amidotransferase A subunit and related amidases	COG0154
R	cysS	553834-555219	Cysteinyl-tRNA synthetase	COG0215
R	tufB	4173523-4174707	GTPases - translation elongation factors	COG0050
R	glnS	705316-706980	Glutamyl- and glutaminyl-tRNA synthetases	COG0008
R	pheT	1793581-1795968	Phenylalanyl-tRNA synthetase beta subunit	COG0072
R	rplK	4176025-4176453	Ribosomal protein L11	COG0080

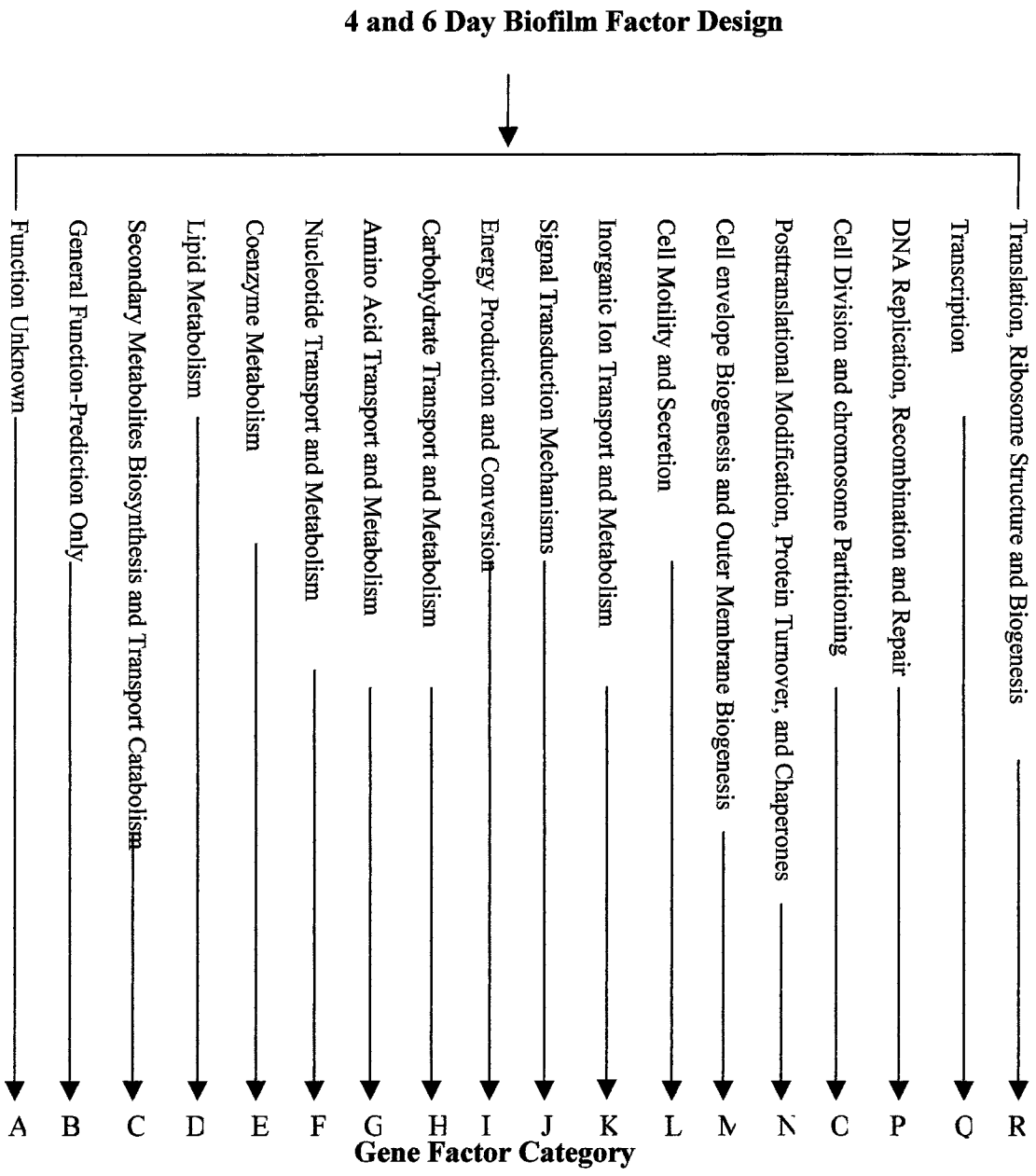


Figure 1. Experimental design for ANOVA. Gene factor is nested in gene function which is crossed and replicated with biofilm maturity of 4 and 6 days. Genes for each gene factor category are described in Table 3. Gene functional categories and genes were extrapolated from NCBI COGs Functional Annotation of *Escherichia coli* and can be reviewed at the following web address: http://www.ncbi.nlm.nih.gov/COG/old/palox.cgi?fun=all.

APPENDIX II – CHAPTER II TABLES AND FIGURES

Table 1. Amino acid sequences for genes of interest within frames expressing respective protein product and homology to known proteins.

Gene and Annotation	Size (aa)	Frame	Amino Acid Sequence
b1995 2066630-2067049	140	+1	LYFVLMRDEMLYNIPCRIYILSTLSLCISGIVSTATATSSETKISNEET LVVTTNRSASNLWESPATIQVIDQQTLQNSTNASIADNLQDIPGVEI TDNSLAGRKQIRIRGWICPYISRHLLSLNPLQARCRRYSRGER*
b2852 2990118-2991492	459	+1	MDLENKFSYHFLEGLTLTEDGILTQGNEQVYIPQKELGVLIVLLESA GHVVLKDMIIESVWKNIIVSDESLTRCIYSLRCIFEKIGYDRCIETIYR KGYRFSGQVFKTKINEDNTSDYSIAIFPFTTSLNTLDPLILNQELVQII SNKKIDGLYTYPMAATNFCNDHISQNSFLSRFKPDYFVTGRINQNN AVNTLYIELIDAKNLFLIASNHLPVDELHNTSQFIIDNILQTVHKPER SVRLAKQDQGYKNHYLSDEMLAGKKELYDFTPESIYRAMTIFDRL QNKSDIQTLKTECYCLLAECHMSLALHGKSELELAAQKALELLDY VSDITTVDGKILAIMGLITGLSGQAKVSHILFEQAKIHSTDIASLYYY RALVHFHNEKIEEARICIDKSLQLEPRRRKAVVIKECVDMYVPNPLK NNIKLYYKETESESHRVIIDNILKLKQLTRICMR*
b1394 1456288-1457076	263	+1	MMEFILSHVEKGVMTLTLNRPERLNSFNDEMHAQLAECLKQVERD DTIRCLLLTGAGRGFCAGQDLNDRNVDPTGPAPDLGMSVERFYNP LVRRLAKLPKPVICAVNGVAAGAGATLALGGDIVIAARSAKFVMA FSKLGLIPDCGGTWLLPRVAGRARAMGLALLGNQLSAEQAHEWG MIWQVVDDETLADTAQQLARHLATQPTFGLGLIKQAINSAETNTLD TQLDLERDYQRLAGRSADYREGVSAFLAKRSPQFTGK*

Table 2. Locations of conserved domain significant alignments where location denotes the amino acid location within the 5,000bp up and downstream sequences.

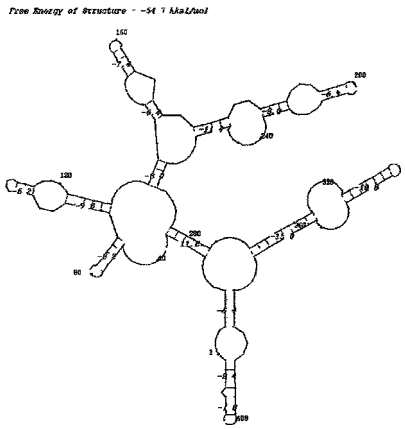
Gene	Conserved Domain	Location	Identity
b1995	CadC	1952-2098	100%
	SLT	2593-2719	100%
b1394	CaiD	1413-1664	100%
	PaaJ	2545-2945	100%
	PaaA, PaaC	235-513	99.6%
	COG3396	234-499	100%
	Thiolase	2545-2813	100%
	Thiolase _C	2819-2944	99.2%
	Pfam01883 DUF59	899-973	100%
b2852	CadC	1-147 (of gene aa sequence)	100%
	Trans_reg_C	27-101 (of gene aa sequence)	100%

Table 3. Locations of significant protein alignments where location denotes the amino acid location within the gene sequences.

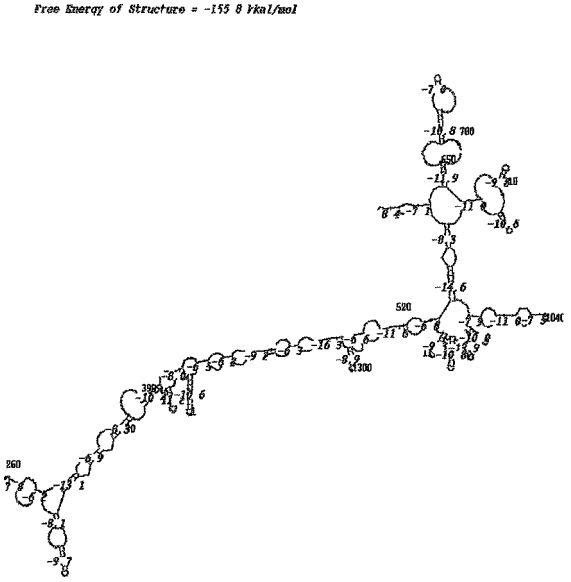
Gene	Protein	Location on Gene amino Acid Sequence	Alignment
b1995 RID: 1051229164-01433-30483	Hypothetical protein probable membrane	1-139	100%
	Hemin receptor precursor in *E. coli*	6-139	100%
	Putative outer membrane receptor for iron or colicin in *E. coli* O157:H7	6-111	100%
	Unknown	6-111	100%
	Hypothetical protein in O157:H7	10-111	100%
	Putative outer membrane receptor for iron or colicin in *E. coli* CFT073	10-111	100%
b2852 RID: 1051230503-011029-1203	Putative invasion protein for *E. coli* K-12	1-458	97%
	ORF o458	1-458	96%
	Putative invasion protein in *E. coli* O157:H7	1-458	94%

Figure 1. RNA secondary structure prediction provided by GeneBee, with A coinciding to b1995, B to b2852, and C to b1394

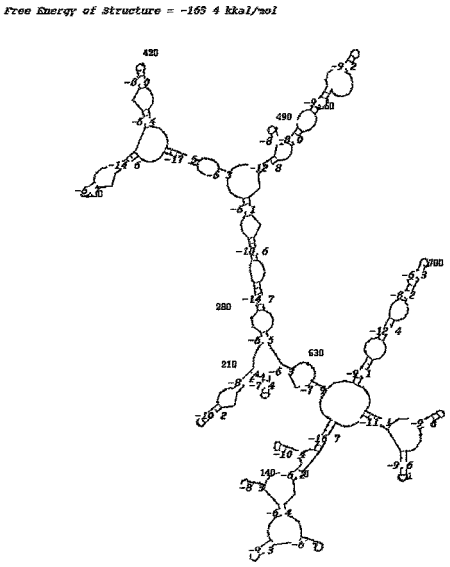
A



B



C



APPENDIX III – CHAPTER III TABLES AND FIGURES

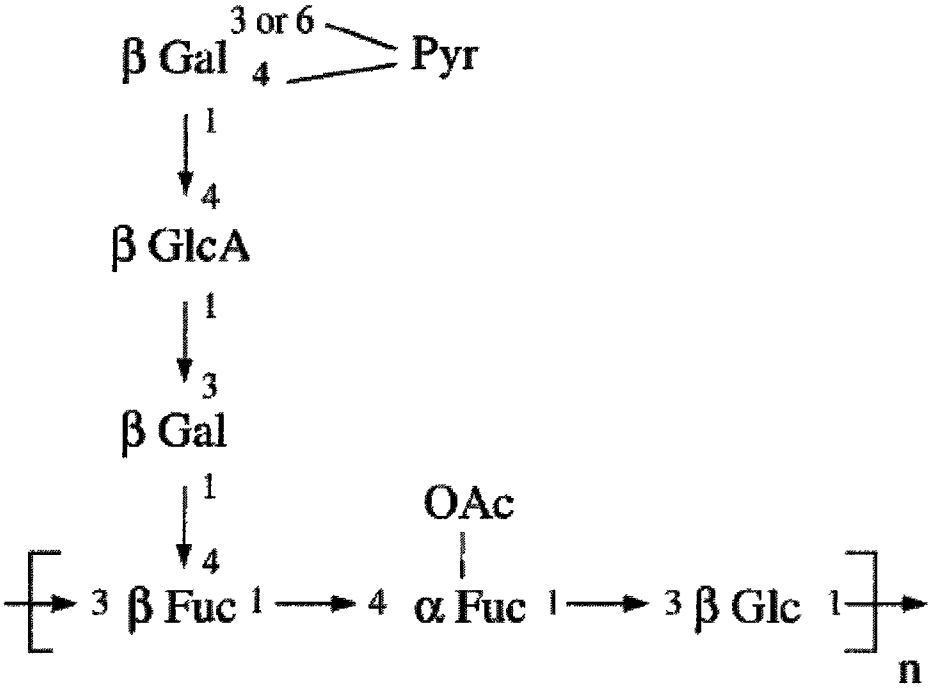


Figure 1. Structure of colanic acid produced in *Escherichia coli* K-12. Abbreviations are Fuc, L-fucose; Gal, D-galactose; GlcA, D-glucuronic acid; Glc, D-glucose; Oac, O-acetyl; Pyr, pyruvate linked acetalically to galactose. Graphic is extrapolated from Stevenson *et al.* (STEVENSON *et al.* 1996) with permission granted by ASM and corresponding author.

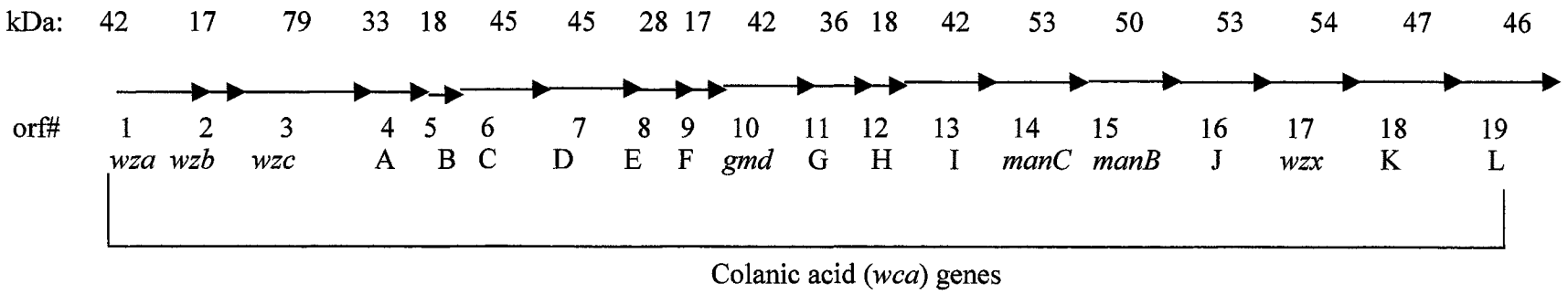


Figure 2. Organization of the colanic acid, *wca*, genes involved with colanic acid production. The *wca* genes are considered an operon and transcription results in a single product. Genes, *manB, manC,* and *gmd* function in the GDP-$_{L}$-fucose biosynthetic pathway. Genes, *wza, wzb, wzc,* and *wzx* have a general role in EPS production. Other genes hypothesized to have a function in colanic acid production are given *wca* names (A to L). Other genes shown are proposed to have no role in colanic acid production. Graphic is extrapolated from Stevenson *et al.* with permission granted by ASM and corresponding author.

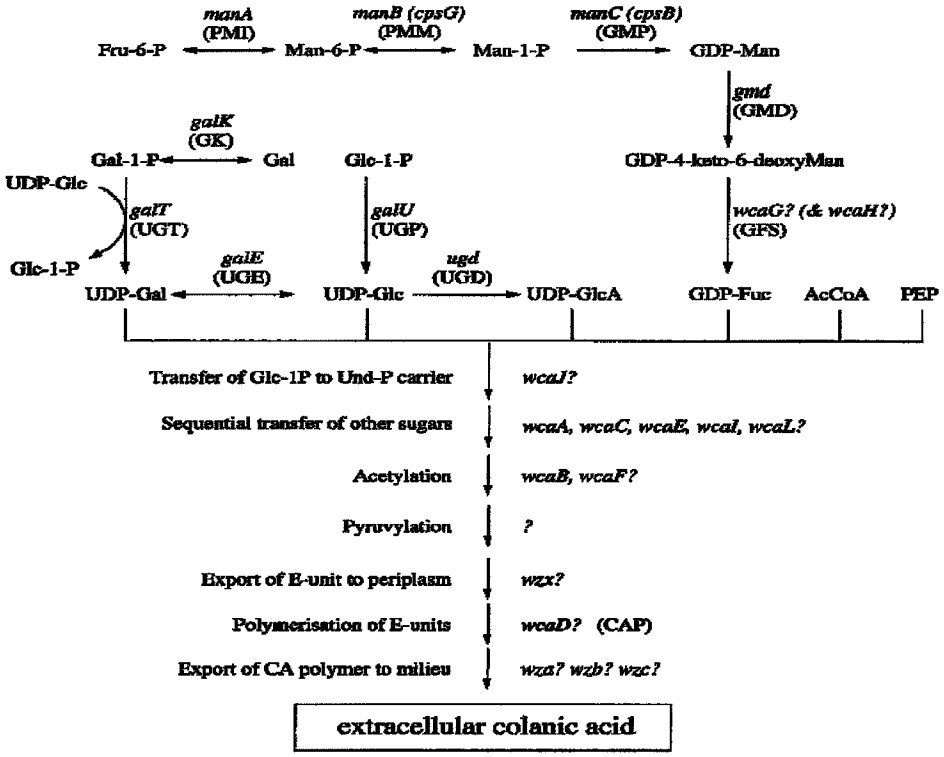


Figure 3. Biosynthesis pathway for extracellular colanic acid. Where AcCoA is acetyl coenzyme A; CAP, colanic acid polymerase; Fru, fructose; GFS, GDP-$_{L}$-fucose synthetase; GK, galactokinase; GMP, GDP-$_{D}$-mannose pyrophosphorylase; GMD, GDP-$_{D}$-mannose dehydratase; Man, $_{D}$-mannose; PEP, phosphoenolpyruvate; PGI, phosphoglucose isomerase; PGM, phosphoglucomutase; PMI, phosphomannose isomerase; PMM, phosphomannomutase; UGD, UDP-$_{D}$-glucose dehydrogenase; UGE, galactose-4-epimerase; UGP, UDP-$_{D}$-glucose pyrophosphorylase; UGT, UTP-$_{D}$-galactose-1-phosphate uridylyl transferase. Question marks indicate hypothesized roles in colanic acid biosynthesis. Graphic extrapolated from Stevenson *et al.* (STEVENSON *et al.* 1996) with permission granted by ASM and corresponding author.

Table 1. New strain designations after transformation of *Escherichia coli* via electroporation for differed expression of colanic acid gene *wzc*. All strains and plasmids were supplied by Christophe Grangeasse, Universitè de Lyon.

Strain Name	Description	Parental Strain	Selection
CLB108	JM83wzc::Km pUC18wzc-rcsA$^+$	JM83wzc::Km	100ug Ampicillin 25ug Kanamycin
CLB109	JM83wzc::Km pUC18wzc569-rcsA$^+$	JM83wzc::Km	100ug Ampicillin 25ug Kanamycin
CLB110	JM83wzc::Km pUC18wzcL6-rcsA$^+$	JM83wzc::Km	100ug Ampicillin 25ug Kanamycin
CLB111	JM83wzc::Km pUC18wzc569L6-rcsA$^+$	JM83wzc::Km	100ug Ampicillin 25ug Kanamycin
CLB112	JM83wzc::Km pUC18wzcK540-rcsA$^+$	JM83wzc::Km	100ug Ampicillin 25ug Kanamycin
CLB113	JM83 pUC18wzc569-rcsA$^+$	JM83	100ug Ampicillin
CLB114	JM83 pUC18wzcL6-rcsA$^+$	JM83	100ug Ampicillin
CLB115	JM83 pUC18wzc569L6-rcsA$^+$	JM83	100ug Ampicillin
CLB116	JM83 pUC18wzcK540-rcsA$^+$	JM83	100ug Ampicillin
CLB117	JM83wzc::Km pUC18-rcsA$^+$	JM83wzc::Km	100ug Ampicillin 25ug Kanamycin
CLB118	JM83 pUC18-rcsA$^+$	JM83	100ug Ampicillin
CLB119	JM83 pUC18wzc-rcsA$^+$	JM83	100ug Ampicillin
JM83wzc::Km	JM83wzc::Km with no plasmid	JM83wzc::Km	25ug Kanamycin
JM83	JM83 with no plasmid F$^-$ *ara glpR208* Δ(*lac-proAB*) *rpsL*(Str^r) [Φ80*dlac*Δ(*lacZ*)M15] *thi*	JM83	No selection

Table 2. Table conveys means for culture and biofilm optical densities of *Escherichia coli* and mutant strains for colanic acid gene *wzc*. Biofilm optical densities were transformed by addition of 1 to all values to meet assumptions of normality for ANOVA and multiple regression analyses. All replications equal 10 unless otherwise specified. Optical densities were taken after 48 hours of growth at 37°C in a polystyrene, 96-well microtiter plate at 570nm. Biofilm optical densities were taken after staining for 15 minutes with 0.1% w/v crystal violet followed by two washes with Millipore water and resuspension in 100% EtOH. Adjusted biofilm means represent multiple regression means for biofilm OD adjusted for culture density as a covariate (r^2=0.5981).

Strain	Mean Culture OD-LB	Mean Culture OD-M63	Standard Deviation Culture OD-LB	Standard Deviation Culture OD-M63	Mean Biofilm OD-LB	Adjusted Biofilm OD-LB	Mean Biofilm OD-M63	Adjusted Biofilm OD-M63	Standard Deviation Biofilm OD-LB	Standard Deviation Biofilm OD-M63
CLB108	1.78 n=20	1 00 n=20	0.40	0 13	1 11 n=20	1.12	1 03 n=19	1.01	0 09	0.07
CLB109	1 53	0 73	0 18	0.02	1 22	1 18	1.01	1 07	0 07	0 01
CLB110	1 65 n=9	0 59	0.19	0 02	1 15	1 28	1.02	1 17	0 12	0 02
CLB111	1 63 n=9	0 66	0 16	0 02	1 19	1 17	1 05	1 05	0 06	0 02
CLB112	1 41	0.73	0 08	0 01	1 26	1.22	1 03	1 11	0 06	0 02
CLB113	1.69 n=19	0.80 n=20	0.49	0.37	1 11 n=20	1 12	1.02 n=20	1 01	0 11	0 02
CLB114	1 72	0 50	0.33	0.05	1.20	1.17	1.02	1.06	0.12	0 01
CLB115	1 45	0 52	0.08	0 05	1 17	1.16	1.03	1 05	0 07	0 01
CLB116	1 44	0 50	0.15	0 04	1 18	1 17	1.04	1 06	0 07	0 01
CLB117	2.10 n=9	0 84	0 55	0 01	1 03	1 08	1 00	0 96	0 01	0 01
CLB118	1 88 n=9	1 05	0 42	0.04	1 16	1 20	1 10	1 09	0 08	0.02
CLB119	1 98 n=9	1 09	0 49	0 11	1 01	1 09	1 01	0 98	0 01	0 00
JM83	1 96 n=15	0 85 n=20	0.42	0 26	1 11 n=20	1 15	1.08 n=20	1 04	0 11	0 08
JM83wzc Km	1 55 n=9	0 82 n=20	0 02	0.23	1.09 n=20	1 12	1 06 n=20	1 01	0 09	0 07

Table 3. Optical densities of biological extracts (presumably colanic acid, non-dialyzable methyl-pentose) from *Escherichia coli* and mutants for colanic acid gene *wzc* after 48 hours of growth at 37°C with shaking at 200rpm. NA values represent optical densities over 4.000, the maximum value readable by instrument. Culture densities measured at 570nm. Spectrophotometer was blanked using trichloroacetic acid subjected to same procedures as cultures. Standards were diluted in Millipore water in same manner as samples. Standards were not extracted from LB, but are listed in column for simplicity. Diluted samples obtained by addition of 900uL of 3% cysteine hydrochloride to 50uL of sample.

Strain	Culture OD-LB	Culture OD-M63	Non-dilute Biological Extract OD-LB (colanic acid) (Abs=396/427)	Non-dilute Biological Extract OD-M63 (colanic acid) (Abs=396/427)	Dilute Biological Extract OD-LB (colanic acid) (Abs=396/427)	Dilute Biological Extract OD-M63 (colanic acid) (Abs=396/427)
CLB108	3.48	3.10	0.34/0.40	0.91/1.11	0.07/0.06	0.04/0.03
CLB109	3.33	2.50	NA/NA	0.03/0.07	0.17/0.13	0.00/-0.00
CLB110	3.34	3.05	0.30/0.37	-0.10/-0.06	0.03/0.03	0.03/0.03
CLB111	3.00	2.31	NA/NA	0.15/0.22	0.37/0.28	-0.00/-0.00
CLB112	3.59	2.93	NA/NA	0.33/0.41	0.23/0.17	0.00/-0.01
CLB113	3.20	3.04	NA/NA	0.16/0.24	0.28/0.21	0.02/0.01
CLB114	3.55	2.64	NA/NA	NA/NA	0.55/0.43	0.15/0.10
CLB115	3.44	2.29	NA/NA	NA/NA	0.59/0.45	0.31/0.29
CLB116	3.31	2.32	NA/NA	NA/NA	0.63/0.48	0.26/0.19
CLB117	3.41	2.51	NA/NA	2.22/2.69	0.24/0.18	0.07/0.05
CLB118	3.59	2.21	0.18/0.27	NA/NA	0.03/0.02	0.27/0.20
CLB119	3.18	2.52	1.82/2.10	-0.06/-0.03	0.07/0.05	0.01/0.01
JM83	3.38	2.43	NA/NA	-0.09-0.05	0.13/0.10	0.03/0.02
JM83wzc::Km	3.44	2.46	NA/NA	-0.05/-0.02	0.19/0.14	0.01/0.00
Neg. Control	0.00	0.00	0.14/0.20	-0.07/-0.04	0.04/0.03	0.03/0.03
5mg/mL Fucose	NA	NA	NA/NA	NA	0.39/0.23	NA
15mg/mL Fucose	NA	NA	NA/NA	NA	0.89/0.54	NA
25mg/mL Fucose	NA	NA	NA/NA	NA	1.87/1.20	NA
35mg/mL Fucose	NA	NA	NA/NA	NA	2.71/1.74	NA
45mg/mL Fucose	NA	NA	NA/NA	NA	NA/2.45	NA
55mg/mL Fucose	NA	NA	NA/NA	NA	NA/2.68	NA
65mg/mL Fucose	NA	NA	NA/NA	NA	NA/3.57	NA
75mg/mL Fucose	NA	NA	NA/NA	NA	NA/NA	NA
85mg/mL Fucose	NA	NA	NA/NA	NA	NA/NA	NA
95mg/mL Fucose	NA	NA	NA/NA	NA	NA/NA	NA

Table 4. Results of 2 factor ANCOVA with CultureOD of *Escherichia coli* and mutants of colanic acid gene *wzc* as the covariate with type III sums of squares. Biofilm optical density (transformed by addition of 1 to all values) used as response variable. Significant differences found between medias, strains, strains within medias, and culture optical densities within strains.

	Df	**Sum of Squares**	**Means Squares**	**F Value**	**Pr (F)**
Media	1	0.0404099	0.04040992	20.90441	0.0000071
Strain	13	0.1389912	0.01069163	5.53088	0.0000000
CultureOD	1	0.0043455	0.00434552	2.24798	0.1348622
Media:strain	13	0.0794997	0.00611536	3.16353	0.0001856
Media:cultureOD	1	0.0017699	0.00176991	0.91559	0.3394216
Strain:cultureOD	13	0.1341525	0.01031942	5.33833	0.0000000
Residuals	294	0.5683257	0.00193308		

VITA

Christa Lynn Bates was born in Springfield, Illinois, and moved to Texas at 12 years of age. She attended Austin Community College in Austin, Texas, for 2 ½ years and finished her Bachelor of Science in Microbiology at Texas State University-San Marcos in May of 2002. As an undergraduate Christa performed independent research resulting in a manuscript in submission and the George H. Meyer award for excellence in microbiology. She also was employed as a laboratory preparation employee for the microbiology labs and taught a microbiology lab her last semester of undergraduate education. Upon graduation, Christa promptly began her graduate education with Dr. Robert J.C. McLean and implemented her thesis design, which was drafted by her as an undergraduate. She also obtained a highly reputable travel grant award from American Society of Microbiology (ASM) in May of 2004. During her graduate career she worked as laboratory instructional assistant for introductory microbiology, pathogenic microbiology, microbial ecology, and genetics.

Permanent Address:
Christa Lynn Bates
1215 West Slaughter Lane, #1913
Austin, Texas 78748

This thesis was typed by Christa Lynn Bates.