12-2013

The Development and Applications of NMR Metabolomics Analysis of Bacterial Metabolomes

Steven M. Halouska
University of Nebraska-Lincoln, halouska@huskers.unl.edu

THE DEVELOPMENT AND APPLICATIONS OF NMR METABOLOMICS ANALYSIS OF BACTERIAL METABOLOMES

by

Steven M. Halouska

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Chemistry

Under the Supervision of Professor Robert Powers

Lincoln, Nebraska

December, 2013

THE DEVELOPMENT AND APPLICATIONS OF NMR METABOLOMICS ANALYSIS OF BACTERIAL METABOLOMES

Steven M. Halouska, Ph.D.

University of Nebraska, 2013

Adviser: Robert Powers

Metabolomics is a relatively new field that involves the study of metabolic responses that are occurring within a biological system. Metabolite profiles of an organism, tissue extract, and biofluids are important indicators to determine the physiological state of a biological profile. Comparison of such profiles from different phenotypes can be used to identify specific metabolic changes leading to the understanding of metabolic pathways, disease progression, drug toxicity and efficacy, and cellular responses to different intracellular and extracellular conditions. Metabolomics investigations often use sophisticated analytical techniques such as NMR spectroscopy to provide an unbiased and comprehensive approach to evaluate metabolic perturbation in different cell lysates.

This dissertation will focus on the development and applications of NMR-based metabolomics methodologies to generate reliable and reproducible results. The protocol has been expanded greatly, optimizing all aspects of the metabolomics process including cell growth, sample preparation, sample handling, data collection, data processing, and data analysis. There are two main approaches in the protocol to decipher NMR metabolomics data: pattern recognition, such as PCA and OPLS-DA, comparing numerous 1-dimensional ^{1}H NMR datasets to analyze biofluids at a global scale, and quantitative profiling of ^{13}C-labeled metabolites using 2-dimensional ^{1}H-^{13}C HSQC. As a

result, our protocol provides a comprehensive analysis, describing unique characteristics and relationships between various samples that differ in their source or treatment.

We applied our protocol to predict the *in vivo* mechanism of action for drug leads from NMR metabolomics data. The NMR analysis resulted in distinct clustering which would be classified by an *in vivo* mechanism. Also, we demonstrated the similarity of *Staphylococcus epidermidis* metabolomes resulting from exposure to divergent environmental stressors that are known to facilitate biofilm formation. Our results suggest that the tricarboxlic acid cycle acts as a metabolic signaling mechanism for the activation of biofilm formation. Also investigated was the mechanism of action of D-cycloserine in *M. smegmatis* and *M. tuberculosis*. Our findings proved that D-alanine-D-alanine ligase is the primary target as cell growth is inhibited when the production of D-alanyl-D-alanine is halted. Furthermore, we were able to identify an alternate path for the production of D-alanine via a possible transaminase mechanism.

AKNOWLEDGEMENT

As I first arrived in the University of Nebraska to pursue a career in science, little did I know what challenges woud lie ahead. The process required a lot of planning, multitasking, and efficient execution to stay on task with the fast progression of the field at hand. Looking back now, I can see that I have made huge progress and matured as a critical thinking individual that is able develop new ideas to further progress the field of science. However, this thesis would have not been completed without the help and support of my advisor, lab mates, collaborators, and the faculty working in the University of Nebraska.

First, I would like to thank my research advisor, Dr. Robert Powers, for his skillful guidance and invaluable help to the development of this thesis. He has taught me valuable lessons that extend even further than the research that occurs in the laboratory settings. I am very grateful for having him as my advisor, because he allowed me develop new ideas and apply them to my research. He was always there for me when I was stuck on different projects. We were able to brainstorm new ideas to move the projects alone. I would like to thank the rest of my committee, Dr. David Berkowitz, Dr. Mark Griep, Dr. David Hage, Dr. Donald Becker, and Dr. Raul Barletta for their time and guidance.

It has been a pleasure working with some very talented colleagues here in the University of Nebraska-Lincoln. I would like to thank Kelly Mercier, Matt Shortridge, Jaime Stark, Jennifer Copeland, Bo Zhang, Bradley Worley, Teklab Gebregiworgis, Shulei Lei, Darrell Marshall, and Jonathan Catazaro. I have enjoyed working in the lab

and going to conferences with you guys. We all had to endure the long group meetings together on a weekly basis. However, those were the most valuable times to understand each other research and create new ideas to further progress our projects. I would also like to thank my very enthusiastic undergraduate, Emily Snell, in our endeavors in developing a sound metabolomics protocol. It's been a pleasure working with her, even though we had to endure the never ending battles with washing NMR tubes.

The development of this thesis is not the work of just one person. Instead the work presented was a result the large contribution of the collaboration of multiple different research groups working together to solve scientific problems. I would like to thank Dr. Raul Barletta for helping me greatly in understanding new biological techniques associated with *Mycobacterium tuberculosis*. It has been a huge pleasure talking about the projects in great details and solving critical problems relating to the field. I also had the pleasure working with Robert Fenton and Denise Zinniel in Dr. Barletta's lab. I would also like to thank Dr. Greg Somerville, Dr. Marat Sadykov, and Dr. Paul Fey for working with them to understand the mechanism of biofilm formation.

Finally I would like to thank my parents and my brothers, Max, Cindy, Scott, and Shane Halouska for their constant support. I should finally be able to give them a good answer when they ask me this question, "Are you done yet?" I would especially like to thank my wife, Dr. Kristin Halouska for her love and support. We both constantly helped each other out when one is in need.

PREFACE

The present dissertation is a collection of published manuscripts (Chapters 3-8) in peer reviewed journals. Chapter 9 is in the process of being submitted, and Chapter 10 has been submitted for further review. The work present in this dissertation has been a highly collaborative process where each of the authors mentioned below provided significant contributions. My contribution to each of the chapters was to develop a sound NMR metabolomics protocol and apply these methods to monitor, compare, and quantify the metabolic perturbations in different organisms or strains under different extracellular conditions. Bo Zhang also contributed substantially in the development and application of the NMR metabolomics protocol, and his contributions should not go unnoticed. Professor Mark Werth and Bradley Worley developed PCA2Tree and PCA utilities software (http://bionmr.unl.edu/pca-utils.php) which provided huge benefits in the NMR metabolomics analysis.

CHAPTER 3:
Revisiting Protocols for the NMR Analysis of Bacterial Metabolomes

Halouska, S[§]., Zhang, B[§]., Gaupp, R., Lei, S., Snell, E., Fenton, R. J., Barletta, R. G., Somerville, G., Powers, R. (2013). Revisiting Protocols for the NMR Analysis of Bacterial Metabolomes. *Journal of Integrated Omics*. Reprinted with permission, copyright Proteomass 2013.

[§] The authors provided equal contributions

CHAPTER 4:
Predicting the *in vivo* Mechanism of Action for Drug Leads Using NMR Metabolomics

Halouska, S., Fenton, R. J., Barletta, R. G., & Powers, R. (2012). Predicting the *in vivo* Mechanism of Action for Drug Leads Using NMR Metabolomics. *ACS Chemical Biology*, *7*(1), 166–171. Reprinted with permission, copyright 2011 by American Chemical Society.

CHAPTER 5:
Tricarboxylic Acid Cycle-Dependent Regulation of *Staphylococcus epidermidis* Polysaccharide Intercellular Adhesin Synthesis

Sadykov, M. R., Olson, M. E., Halouska, S., Zhu, Y., Fey, P. D., Powers, R., & Somerville, G. A. (2008). Tricarboxylic Acid Cycle-Dependent Regulation of *Staphylococcus epidermidis* Polysaccharide Intercellular Adhesin Synthesis. *Journal of Bacteriology*, *190*(23), 7621–7632. Reprinted with permission, copyright 2008 American Society for Microbiology.

CHAPTER 6:
Using NMR metabolomics to investigate tricarboxylic acid cycle dependent signal transduction in *Staphylococcus epidermidis*

Sadykov, M. R., Zhang, B., Halouska, S., Nelson, J. L., Kreimer, L. W., Zhu, Y., Powers, R., Somerville, G.A. (2010). Using NMR metabolomics to investigate tricarboxylic acid cycle dependent signal transduction in *Staphylococcus epidermidis*. *Journal of Biological Chemistry*, *285*(47), 36616–36624. Reprinted with permission, copyright 2010 American Society for Biochemistry and Molecular Biology.

CHAPTER 7:
NMR analysis of a stress response metabolic signaling network

Zhang, B[§]., Halouska, S[§]., Shiaffo, C. E., Sadykov, M. R., Somerville, G. A., & Powers, R. (2011). NMR analysis of a stress response metabolic signaling network. *Journal of Proteome Research*, *10*(8), 3743–3754. Reprinted with permission, copyright 2011 American Chemical Society.

[§] The authors provided equal contributions

CHAPTER 8:
CcpA Regulates Arginine Biosynthesis in *Staphylococcus aureus* through Repression of Proline Catabolism

Nuxoll, A. S., Halouska, S., Sadykov, M. R., Hanke, M. L., Bayles, K. W., Kielian, T. W., Powers, R., Fey, P.D. (2012). CcpA Regulates Arginine Biosynthesis in *Staphylococcus aureus* through Repression of Proline Catabolism. *PLOS Pathogens*,*8*(11). Reprinted with permission, copyright 2012 PLOS.

CHAPTER 9:
Role of Alanine Racemase in Mycobacterium smegmatis and *Mycobacterium tuberculosis*

Realp, T., Pandey, A., Halouska, S., Fenton, R. J., Zinniel, D. K., Powers, R., Barletta, R. G., Robledo, J., Chacon, O. Role of Alanine Racemase in *Mycobacterium smegmatis* and *Mycobacterium tuberculosis*.

CHAPTER 10:
Insights into the Mechanism of Action of D-cycloserine in Mycobacteria

Halouska, S., Fenton, R. J., Zinniel, D. K., Marshall, D. D., Chacon, O., Barletta, R. G., & Powers, R. Insights into the Mechanism of Action of D-cycloserine in Mycobacteria.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Introduction

Cells are complex systems involving numerous chemical, electrochemical, and mechanical processes which govern basic cellular activities and coordinate cell actions. It was once thought that the many processes occurring in cells display a hierarchal level of organization, transferring information from DNA to RNA to proteins [1-3]. However, a cell's behavior is a result of the complex interactions and regulations between these chemical entities (e.g., DNA, RNA, proteins, small molecules). For example, transcription factors, which are products of gene transcription, can inhibit or activate the transcription of mRNAs [4-7]. Proteins can interact with other proteins, forming new and larger complexes with different functions [8, 9]. Biochemical reactions in cellular metabolism can be integrated as large networks that are regulated by enzymes. Likewise, small molecule interactions can also regulate gene transcription and protein function [10-12]. It is important to note that biomolecular interactions and regulations are not at all static. Instead, they change dynamically in response to changes in the environment and intercellular state. The main aspect of these functional responses concerns perturbations in metabolic regulations where enzymes, which catalyze metabolite conversions, are subjected to multiple levels of regulations [13]. Also, transcriptional regulations may play important roles in metabolic responses [13]. This is where using genomic,

proteomic, and metabolomics approaches are beneficial to understand biological behavior with different conditional perturbations (Figure 1.1a) [14-18].

A key aspect to understanding a biological system is to study metabolism, as the concentration of the metabolites is sensitive to environmental and genetic changes [19]. Furthermore, a complete set of metabolites, which are products (molecular weight, less than 1,500 daltons) of enzyme-catalyzed reactions in a cell, is referred to as the metabolome [20-22]. A comprehensive list of metabolites can be used to generate multiple metabolic networks (e.g., Glycolysis, TCA cycle, urea cycle, etc.), which can be visualized by databases such as KEGG and Metacyc [23, 24]. Since many metabolites are found in a wide range of organisms while others are family or species specific, it is likely that some metabolites will fulfill important biological functions [20]. Hence, analyzing the metabolome is well suited to determine perturbed network patterns or metabolite changes to address important biomarkers for specific diseases, mutation, or determining the function of drug candidates (Figure 1.1b) [18]. To understand the different set of conditions perturbing a biological system, a comprehensive analysis in which all metabolites are identified and quantified is performed. This comprehensive approach is known as " metabolomics" [20].

Metabolomics is one of the newest disciplines of the –omics cascade (genomics, transcriptomics, proteomics) that revolutionized the strategy of modern research. Hypothesis-based research has remained the golden rule of research [25]. However, devising a hypothesis can be difficult with the ever increasing complexity of a problem.

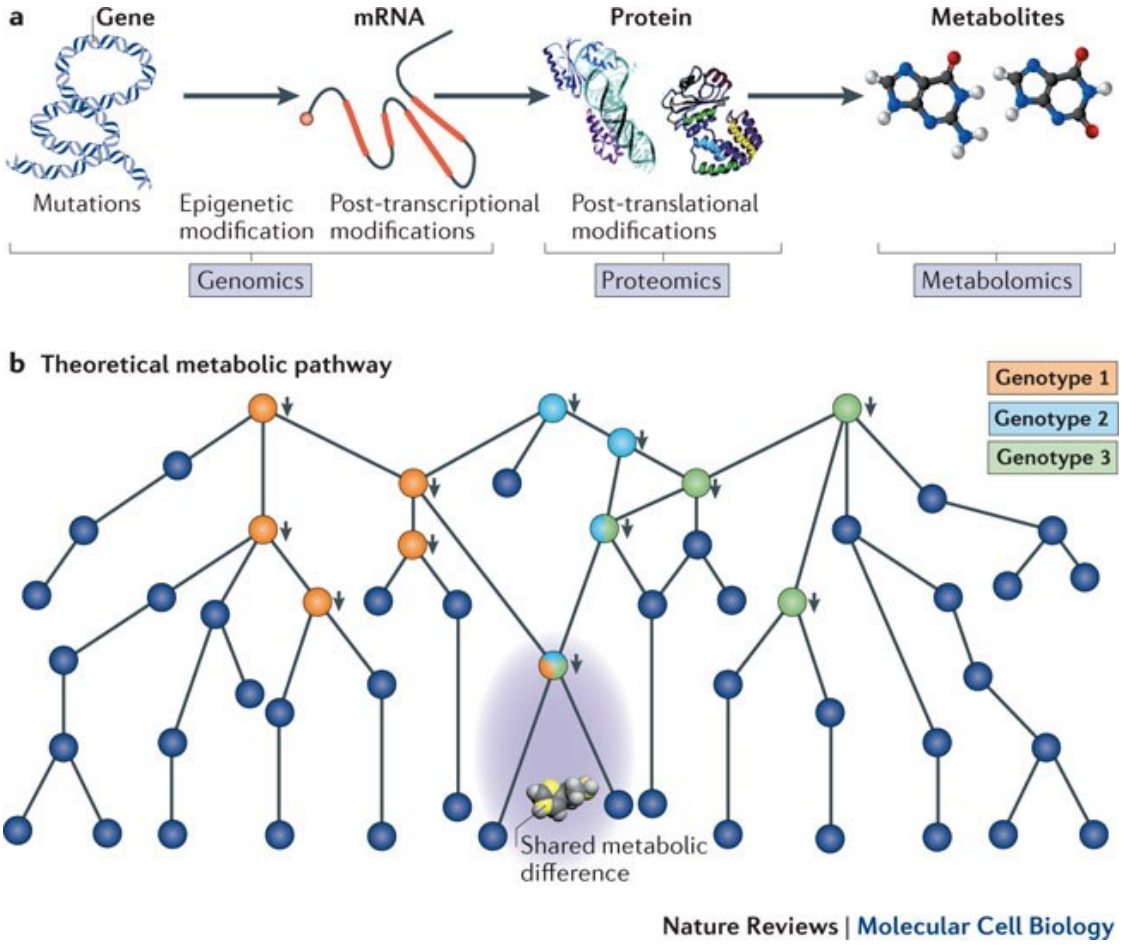


Figure 1.1. a) Schematic describing the downstream processes of the omics cascade. b) A schematic of a metabolic pathway. Metabolites are represented as circles. An alteration in a single enzyme can lead to a cascade of metabolic perturbations. Here, metabolites whose levels are altered in each of three theoretical genotypes are shown by colored circles. Orange, light blue, and green circles represent metabolites altered in genotype 1, genotype 2 and genotype 3, respectively. Metabolites whose levels are similarly altered in multiple genotypes are represented by multi-colored circles. Reprinted with permission from [18] © MacMillan Publishers Limited, 2012.

Instead, metabolomics can be performed without any prior knowledge of the system. This means unique metabolic patterns and perturbed metabolites can be used toformulate a new hypothesis. Hypothesis-based approaches and metabolomics are complementary in that the analysis of both approaches can lead to similar findings, thus strengthening our knowledge in the field [26]. Sometimes metabolomics can address unexpected findings that are not related to the original problems being investigated. This can lead to new areas to explore and a new hypothesis to be developed.

Metabolomics offers collaborative opportunities, connecting biology centric research with high throughput analytical platforms and interpretation while applying computer generated models to answer specific questions or concerns [27]. Collaborative efforts using metabolomics have been advantageous for all aspects of science, where groups of researchers can work together with similar interests from the same or different institution. It also brings together researchers with different viewpoints and approaches to solve a problem. Diversity of experience and skill sets may increase the likelihood of a breakthrough or of successful research. Collaborations have become a popular trend, where national centers and international projects have been developed to understand and characterize different metabolomes [27]. It has been advantageous from a graduate student's perspective as it can expand our horizons in multiple disciplines.

Although, metabolomics has been around for about a decade, it has grown quite rapidly and has become a widely-used approach with a broad field of applications, including plants [28-30], bacteria [31-33], eukaryotic cells [34], medicinal [35], and veterinary sciences [36]. For example, metabolomics research has been heavily used in

the discovery of biomarkers or patterns associated with diseases [37, 38]. Metabolomics has been applied in the study of drug toxicity and efficacy, which can lead to novel and potent drugs [39-41]. Metabolomics, along with the genomics and proteomics has the capability of providing diagnostics and prognostics of drug discovery clinical trials [42]. Instead, metabolomics is able to identify biomarkers that are associated with hepatotoxicity and nephrotoxicity, which are the two major reason most drug candidates fail [43]. New metabolomics technologies can be developed to create non-invasive techniques that can be used to diagnose and monitor disease progression, recognize response to treatments, and identify the development of resistance to treatments.

Metabolomics methodologies usually involve information rich analytical platforms including: nuclear magnetic resonance spectroscopy (NMR); mass spectroscopy (MS) coupled with separation techniques such as gas chromatography (GC) or liquid chromatography (LC); and Fourier transform-infrared spectroscopy (FT-IR) [44]. NMR is a popular tool because it provides a rapid, non-destructive, high-throughput method that requires minimal sample-handling. Typically, metabolites are detectable by ^{1}H-NMR because they usually contain one or more protons in their molecular structure, and each proton produces one or more peaks. However, a metabolite mixture from a given biological system may consist of thousands of metabolites, thereby producing thousands of proton signals in a given spectrum and providing a wealth of information on the nature of the biological system (Figure 1.2).

NMR is somewhat of a universal platform as it can capture a wide range of metabolites with different chemical characteristics, such as aromatic compounds,

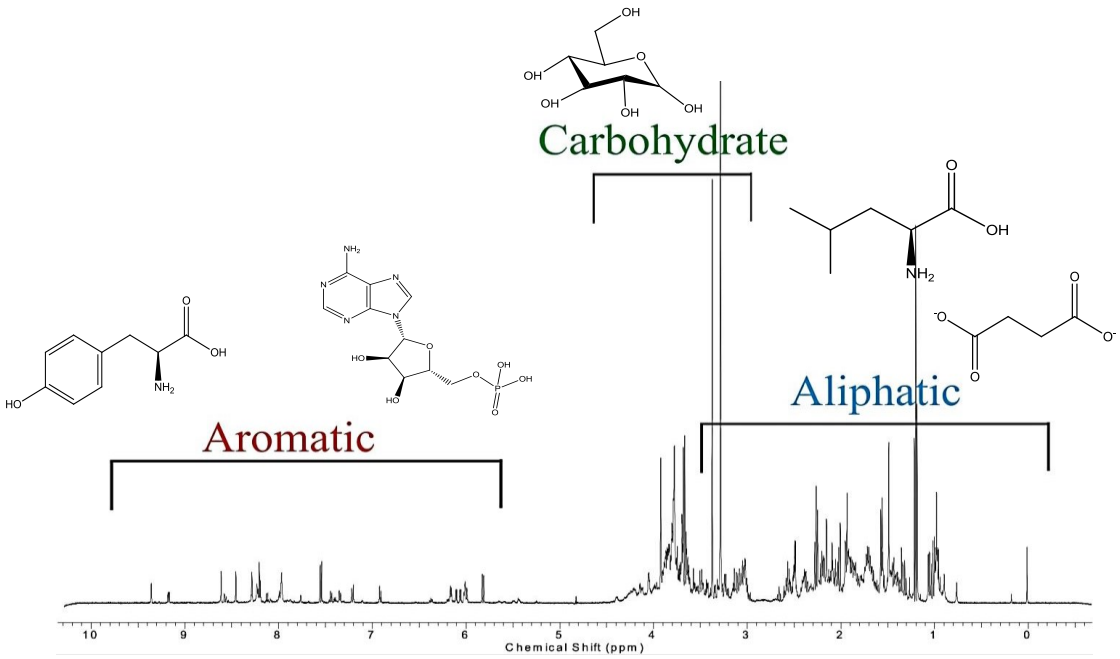


Figure 1.2. A 1D ^{1}H NMR spectrum, displaying hundreds of metabolites found in an *E.coli* metabolomics sample. The NMR spectrum can be divided into three regions describing the type of metabolites: aromatic, carbohydrates, and aliphatics.

carbohydrates, amino acids, and aliphatic compounds. Unfortunately, analysis of every peak can be very challenging due to the limitation of the proton spectral width, leading to peak overlap.

An accurate but qualitative approach involves the use of statistical analysis and pattern recognition. Pattern recognition techniques such as principle component analysis (PCA) and orthogonal projection to latent structure discriminant analysis (OPLS-DA) are currently being employed to capture the global perturbations in a metabolome [45, 46]. PCA is an unsupervised method that reduces a multivariate dataset to a single point and projects the major variations in the dataset into a few axes called principal components. In this way, spectral variations are captured in a model that can be easily visualized. The overall underlying assumption is that similar metabolomic profiles will cluster closely together in PCA or OPLS-DA scores plots and that differing metabolomics profiles will cluster away from each other [47].

OPLS-DA is a supervised method used to determine the variations within the data set that are correlated to classification labels (e.g., treatment, control, disease state) set by the investigators. If the variations within the data set are not correlated with the classification label, then the variations are filtered out, resulting in a single latent vector [46]. This is analogous to PCA because PCA captures all the intrinsic variations within the data set. Since the classification labels are defined in OPLS-DA, the scores plot will create distinct clusters between classes regardless of their similarities. A separation between classes will even occur for randomly generated data [48]. Thus, it is essential that OPLS-DA models are properly validated. However, with proper validation using

both PCA and OPLS-DA methodologies, these tools are extremely powerful tools used to organize and classify different strains, drug treatments, or diseases. This also allows for the determination of the mechanisms of drugs with unknown actions, by comparing the results to those for classes of drugs with known mechanisms of action [49].

A major problem with 1D-NMR metabolomics techniques is the inability to accurately quantitate metabolic perturbation due to overlapping peaks. In addition, attempts to identify these metabolites present a great challenge. Therefore, the use of a two-dimensional (2D) ^{1}H-^{13}C HSQC experiment leads to a dispersion of peaks along the ^{1}H and ^{13}C axes, increasing the resolution of the spectrum (Figure 1.3). More importantly, a 2D ^{1}H-^{13}C HSQC spectrum offers useful information for the identification of individual metabolites, since, multiple peaks comprising coupled ^{1}H and ^{13}C chemical shifts may be attributed to a single metabolite, providing several redundant data points for confirming an assignment. A number of databases have been developed to enable metabolite identification by matching the chemical shifts from a 2D ^{1}H-^{13}C HSQC spectrum obtained for a metabolomics sample to a collection of reference NMR spectra for numerous known metabolites. Such databases include the Human Metabolome Database (HMDB) [50], Madison Metabolomics Database (MMCD) [51], and Platform for Riken Metabolomics (PRIMe) [52]. These databases consist of hundreds to thousands of metabolites from a broad range of organisms; therefore, metabolite identification should be verified with KEGG and Metacyc databases [23, 24, 40]. Metabolites can be quantified using time-zero 2D ^{1}H-^{13}C HSQC [53] or by quantitating

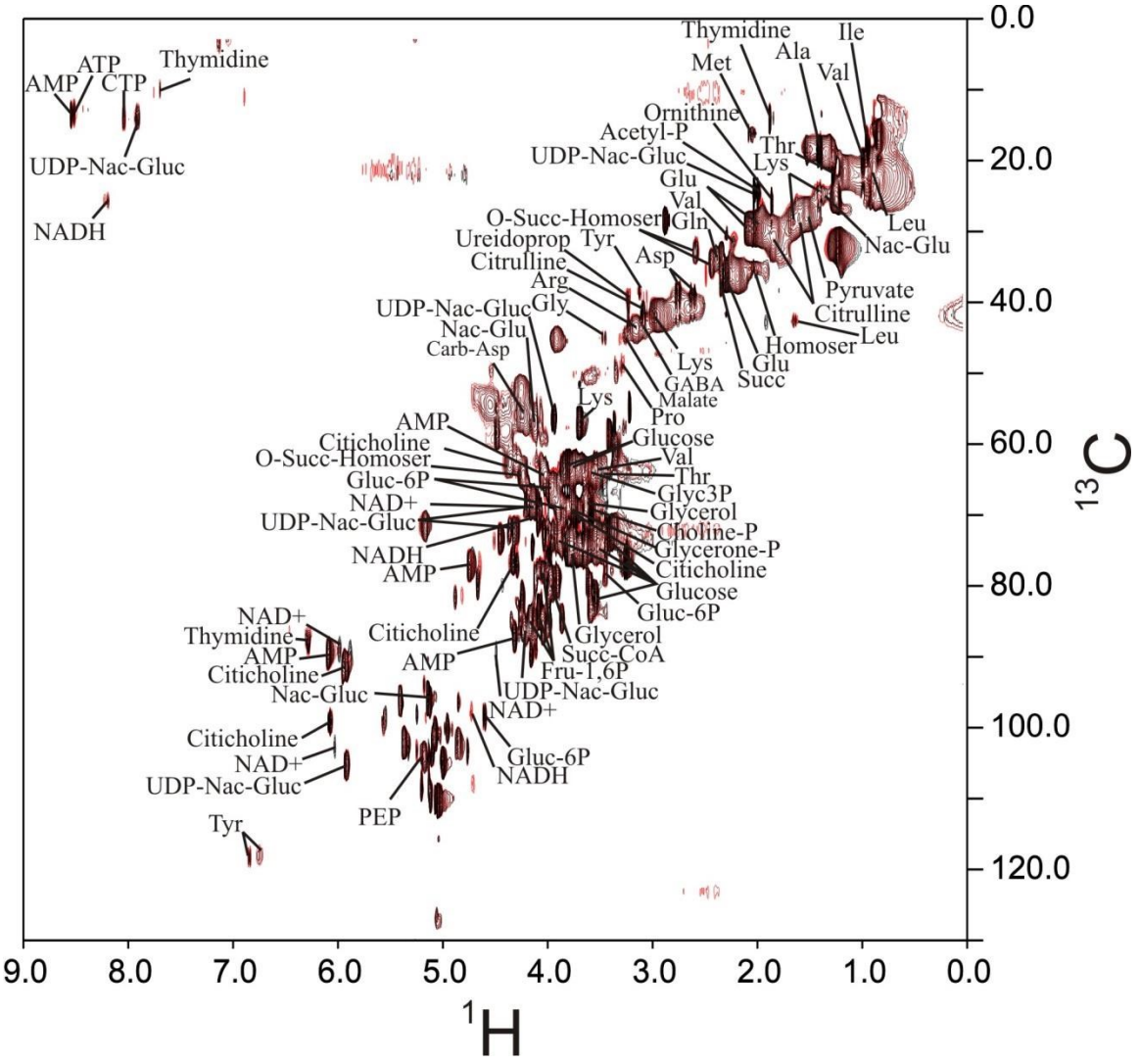


Figure 1.3. An example of a typical 2D ^{1}H-^{13}C HSQC spectrum obtained for a *M. smegmatis* culture grown in MADC media containing ^{13}C-glycerol. The metabolites were identified using HMDB, MMCD, and Prime databases.

the relative difference in peak height [54]. Heat-maps, bar graphs or metabolic pathways can then be generated to compare concentrations or relative peak heights of different metabolites to identifying major perturbations associated with genetic or environmental changes.

1.2 Overview of Thesis

1.2.1 Development of Metabolomics Methodologies

Metabolomics is a relatively new field, but it has evolved into a powerful tool for understanding the global physiological and pathological states of a biological system. This dissertation will focus on the development of NMR-based metabolomics methodologies to ensure that the required information is obtained in a high throughput and reliable manner. These NMR-based metabolomics methods were developed while applying the techniques to different organisms, such as *Aspergillus nidulans*, *Mycobacterium tuberculosis*, *Mycobacterium smegmatis*, *Staphylococcus aureus*, *Staphylococcus epidermidis*; biofluids (urine), and human cell lines (pancreatic cancer cells) were also examined [47, 54-58].

Chapter 2 will address the application of pattern recognition tools, such as PCA and OPLS-DA, which can be used to decipher global changes in large NMR datasets of cell lysates or biofluids. The chapter will explain the importance of the validation process and the development of statistically meaningful plots. Statistical tools such as PCAtoTree [59] and PCA/PLS-DA utilities [60] have been developed to quantify the

differential relationship between clustering patterns. S-Plots are then used to identify metabolite changes that are responsible for class separation.

Chapters 3 outlines the overall metabolomics protocol that was optimized in order to generate reliable and reproducible data. Initially, the protocol was limited to only the use of PCA and the comparison of a control, mutant, and drug treated cell lines [47]. The protocol has been expanded greatly, optimizing all aspects of the metabolomics processes including cell growth, sample preparation, sample handling, data collection, data processing, and data analysis. The protocol now includes the qualitative and quantitative analysis of 2D ^{1}H-^{13}C HSQC spectra. Peak intensities and volumes within a 2D ^{1}H-^{13}C HSQC spectrum are directly proportional to the concentration of a given metabolite. However, conventional HSQCs are difficult to quantify because different spin dynamic properties, J-coupling, translational relaxation, and transverse relaxation for each atom in a metabolite also affect peak intensities and volumes [53]. Therefore, the protocol incorporates two methods: quantifying metabolic perturbations comparing relative peak heights, and quantifying the concentrations of metabolites using time-zero HSQC [53].

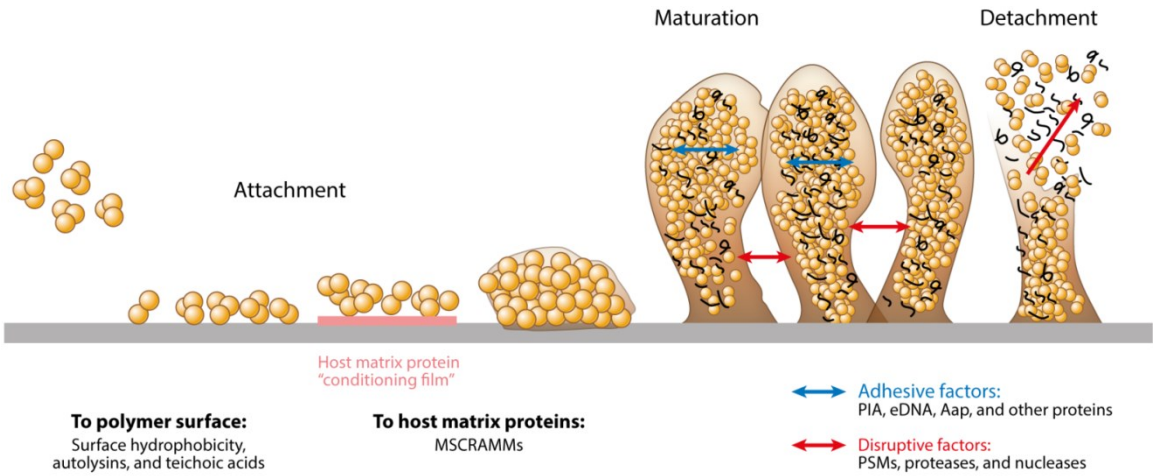
Chapter 4 discusses a new methodology to predict the *in vivo* mechanism of action for drug leads from NMR metabolomics data. Initially, drugs were compared with mutant knockout strains to determine the mechanism of action. Drugs that clustered similar to the mutant strains in PCA or OPLS-DA scores plots were said to have a similar impact on the metabolome as the genetically inactivated protein and, therefore, the genetically inactivated protein was identified as the likely drug target [47]. However, this

requires some general knowledge of the potential drug target in order to select an appropriate mutant knockout strain. Thus, this approach was further developed by comparing chemical leads with unknown mechanisms to reference drugs with known protein targets. We postulated that drugs targeting the same protein would likely cluster together in a PCA or OPLS-DA scores plot. Three chemical leads shown to inhibit *M. tuberculosis* and *M. smegmatis*, but with an unknown mechanism of action, were compared with known mycobacterial drugs. The NMR analysis resulted in multiple drugs with similar mechanisms of action clustering closely together in an OPLS-DA scores plot, allowing for the prediction of a mechanism of action for the three chemical leads with activity against *M. tuberculosis* and *M. smegmatis*.

1.2.2 Biofilm Formation

Staphylococcus epidermidis and *Staphylococcus aureus* are the most prevalent bacteria found colonizing human skin and mucous membranes [61]. Despite their low virulence, they have been the leading causes of infections during medical care [62]. These infections are most frequently a result of cuts, medical implants, dialysis catheters, and many other medical procedures. The virulence and survivability of these bacterial species is associated with their ability to attach to surfaces and form a sticky residue called a biofilm, in which larger colonies of cells can be formed (Figure 1.4a) [63]. At a molecular level, polysaccharide intercellular adhesion (PIA) plays an essential role in bacterial adherence. PIA is primarily composed of a mixture *N*-acetylglucosamine in β-1,6-glycosidic linkages containing deacetylated amino groups, succinate, and phosphate

a)



Otto M. 2013.
Annu. Rev. Med. 64:175–88

b)

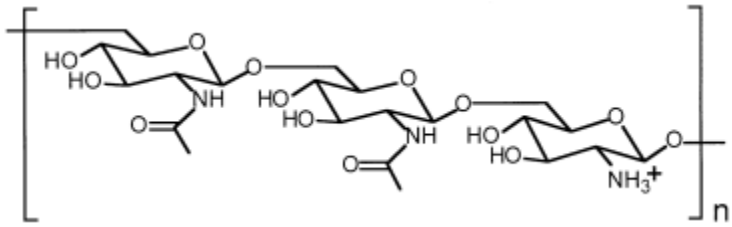


Figure 1.4. a) The process of biofilm formation. The cells first attached to a polymer surface. As more cells are attached to the surface, they secrete out more PIA. The biofilm matrix matures and then overtime, parts of the biofilm detaches forming new colonies. b) Molecular structure of PIA. Reprinted with permission from [63] © Annual Review of Medicine, 2013.

substitutions (Figure 1.4b) [64, 65]. Overtime, more bacterial cells adhere to the complex, secreting more PIA. Channels are formed through the complex which allows for nutrients to flow deep within the complex community [66]. Parts of the complex colony can then break off and form new colonies [63]. Overall, biofilms allow bacterial cells to survive and grow through harsh environmental conditions. Numerous antibiotics are ineffective or have diminished activity against a biofilm, because of the inability to penetrate through the sticky mass [65, 67]. Therefore, studies are being performed to find drugs that can detach biofilms from surfaces or prevent biofilm formation [68].

In this study, we demonstrated for the first time that a metabolic response, TCA cycle repression, can act as a signaling mechanism for the activation of PIA and biofilm formation. In **Chapter 5,** *S. epidermidis* TCA cycle mutants were constructed, and the function of central metabolism was analyzed. The results showed that TCA cycle inactivation altered the metabolic status of *S. epidermidis*, resulting in a derepression of PIA biosynthetic genes and a redirection of carbon from growth into PIA biosynthesis.

In **Chapter 6,** we demonstrate that environmental stressors (iron limitation and addition of ethanol) decrease TCA activity resulting in metabolic changes. These metabolic changes are sensed by a metabolite-response regulator (catabolite control protein; CcpA) that affect PIA production.

Chapter 7 describes a detailed NMR metabolomics analysis to investigate a diverse array of environmental stress factors associated with biofilm formation, such as 5% NaCl, 2% glucose, 0.06 ug/ml tetracyclin, 400 nM autoinducer-2, 4% ethanol, and iron limitation. A detailed quantitative NMR analysis describing the changes in

metabolite concentration using 2D ^{1}H-^{13}C HSQC and 2D ^{1}H-^{1}H TOCSY spectra was used to describe the perturbation of the metabolome associated with the environmental stresses. The results further demonstrate that different environmental stress conditions induce inactivation of the TCA cycle. Moreover, the TCA cycle inactivation plays a central role in the proposed signaling pathway that also involves the perturbation of other metabolites associated with PIA formations. Furthermore, our metabolomics analysis has shown that some environmental stresses such as the addition of NaCl and autoinducer-2 did not affect the TCA cycle demonstrating that these factors may perturb other distant processes.

Chapter 8 illustrates through genetic and biochemical approaches that *S. aureus* ccpA can synthesize arginine from proline via the urea cycle. This is the first report of bacteria utilizing proline for arginine biosynthesis. Ccpa is a known gene regulator for carbon catabolite repression, and can de-repress arginine biosynthesis via glutamate [56, 69]. Ccpa can also repress TCA cycle activity which is important in biofilm formation [70].

1.2.3 Mechanism of action of DCS on Mycobacterium tuberculosis and Mycobacterium smegmatis

D-Cycloserine (DCS) is an effective second line antibiotic used to treat *Mycobacterium tuberculosis*. DCS is a cyclic analogue of D-alanine and interferes with the formation of peptidoglycan biosynthesis by the competitive inhibition of alanine racemase and D-alanyl-D-alanine ligase. Both enzymes are unique to bacteria and

valuable drug discovery targets for developing selective antibacterial agents [71]. However, DCS is rarely used due to serious neurological side effects that result from an interaction with neuronal N-methyl aspartate receptors [72-74]. Therefore, understanding the *in vivo* mechanism of action for DCS and determining if alanine racemase or D-alanine-D-alanine ligase is the lethal target of DCS could aid in the development of effective and safe antitubercular agents.

Although it is known that DCS can competitively bind to both alanine racemase and D-alanine ligase, it is unclear which enzyme is the lethal target of DCS [75]. Recent studies have shown that mutations leading to the overproduction of alanine racemase increased resistance to DCS [76]. Similarly, alanine racemase null mutants are susceptible to DCS [76]. Overall, the sensitivity of DCS to both overproduction strain and null mutants may suggest that alanine racemase plays an important role in the *in vivo* mechanism of DCS. However, the overproduction of D-alanine-D-alanine ligase has been shown to reduce susceptibility of DCS, suggesting that D-alanine-D-alanine ligase may also play an important role in DCS activity [75]. Moreover, DCS mediated inhibition can be reversed by the presence of D-alanine in the medium, suggesting DCS might inhibit growth by alanine racemase alone [77]. However, our NMR metabolomics experiments show that the *in vivo* mechanism of inactivation of alanine racemase by mutation does not match the *in vivo* mechanism of action for DCS (Figure 1.5) [55]. This suggests that another enzyme, possibly D-alanine-D-alanine ligase, may be the lethal target.

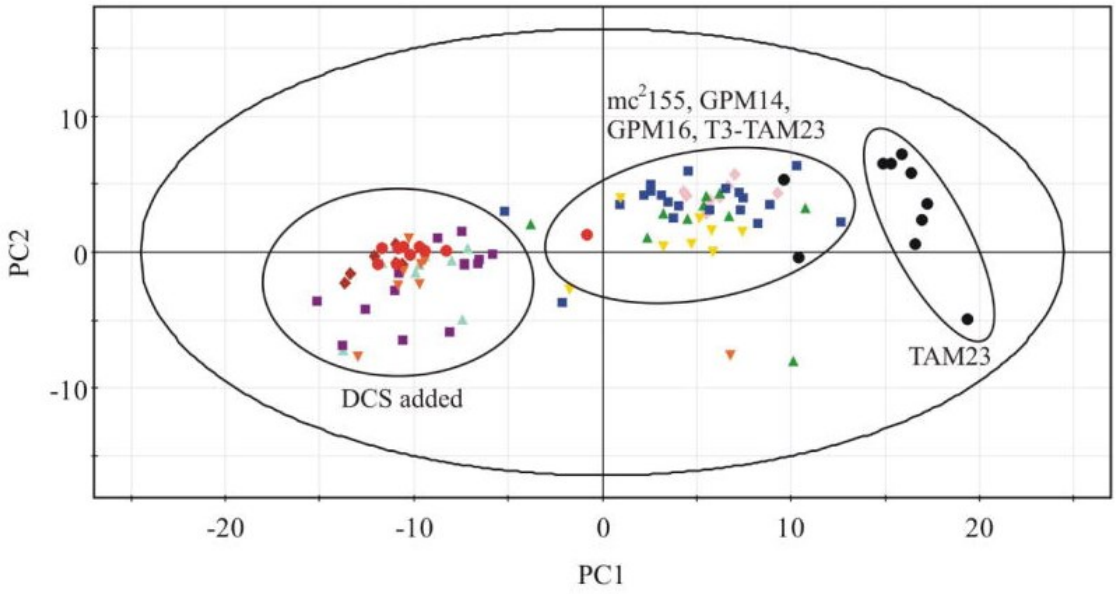


Figure 1.5. a) PCA scores plot comparing wild-type mc^{2}155(■), alanine racemase null mutant (TAM23) (●), alanine racemase over producing mutant (GPM14) (♦), DCS resistant strain unrelated to alanine racemase mutations (GPM 16) (▲), TAM23 complemented with wild-type alanine racemase gene (TAM23 pTAMU3) (▽), mc^{2}155 with DCS (■), TAM23 with DCS (●), GPM14 with DCS (♦), GPM16 with DCS (▲) and TAM23pTAMu3 with DCS (▼). Cultures that are clustered together displays similar metabolomes while distant clustering patterns have different metabolomes. Reprinted with permission [55] © American Chemical Society, 2007.

The essentiality of alanine racemase in mycobacteria is also a point of discussion because it is important in determining whether alanine racemase or D-alanine-D-alanine ligase is the lethal target of DCS [78, 79]. For example, if alanine racemase is essential, then it would be assumed that the inhibition of alanine racemase by DCS would be lethal. Therefore, developing effective and safer antibiotics would focus on alanine racemase as a drug target. On the other hand, if alanine racemase in not essential, meaning there is an alternate path for the production of D-alanine, then D-alanine-D-alanine ligase may assume the role as the potential lethal target of DCS. In our previous study using NMR metabolomics, we were able to verify that there is a potential alternate pathway for the production of D-alanine, where an alanine racemase null mutant, an alanine racemase overproduction mutants, and wild type cells were *all* able to grow in the absence of D-alanine [55]. Furthermore, our 1D ^{1}H NMR metabolomics analysis showed a decrease in glutamate in the alanine racemase null mutant cell lysate samples, suggesting glutamate is being utilized to produce D-alanine via a transaminase mechanism [55].

A previous NMR metabolomics study provided substantial insights regarding the role of alanine racemase to DCS activity [55]. Nevertheless, further studies were needed to determine the essentiality of alanine racemase and for defining the role of D-alanine-D-alanine ligase in DCS activity. **Chapter 9** describes further investigations of the essentiality of alanine racemase to *M. smegmatis*, and provides additional evidence for a possible alternate route for the production of D-alanine. It also illustrates the importance of media selection in cell cultures, since the use of different media may generate contradictory results.

Chapter 10 describes further investigations into the identification of the *M. tuberculosis* and *M. smegmatis in vivo* lethal target of DCS. A detailed NMR metabolomics analysis identified D-alanine-D-alanine ligase as the lethal target of DCS. Inhibition of cell growth by DCS was correlated with a halt in the production of D-alanyl-D-alanine. Also, increasing the cellular pool of D-alanine was identified as a mechanism of DCS resistance. Thus, inhibition of alanine racemase may contribute indirectly to DCS activity by lowering the levels of D-alanine and allowing DCS to outcompete D-alanine for D-alanine-D-alanine ligase binding.

1.3 References

1. Crick, F., *Central Dogma of Molecular Biology.* Nature, 1970. **227**(5258): p. 561-3.

2. Beadle, G.W. and E.L. Tatum, *Genetic Control of Biochemical Reactions in Neurospora.* Proceedings of the National Academy of Sciences of the United States of America, 1941. **27**(11): p. 499-506.

3. Shapiro, J.A., *Revisiting the Central Dogma in the 21st century.* Annals of the New York Academy of Sciences, 2009. **1178**: p. 6-28.

4. Pahl, H.L., *Activators and Target Genes of Rel/NF-kappaB Transcription Factors.* Oncogene, 1999. **18**(49): p. 6853-66.

5. Orphanides, G., T. Lagrange, and D. Reinberg, *The General Transcription Factors of RNA Polymerase II.* Genes & Development, 1996. **10**(21): p. 2657-83.

6. Odom, D.T., et al., *Control of Pancreas and Liver Gene Expression by HNF Transcription Factors.* Science, 2004. **303**(5662): p. 1378-81.

7. Rao, A., C. Luo, and P.G. Hogan, *Transcription Factors of the NFAT Family: Regulation and Function.* Annual Review of Immunology, 1997. **15**: p. 707-47.

8. Ito, T., et al., *Toward a Protein-Protein Interaction Map of the Budding Yeast: A Comprehensive System to Examine Two-Hybrid Interactions in all Possible*

Combinations Between the Yeast Proteins. Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(3): p. 1143-7.

9. Jones, S. and J.M. Thornton, *Principles of Protein-Protein Interactions.* Proceedings of the National Academy of Sciences of the United States of America, 1996. **93**(1): p. 13-20.

10. Howitz, K.T., et al., *Small Molecule Activators of Sirtuins Extend Saccharomyces cerevisiae Lifespan.* Nature, 2003. **425**(6954): p. 191-6.

11. Camilli, A. and B.L. Bassler, *Bacterial Small-Molecule Signaling Pathways.* Science, 2006. **311**(5764): p. 1113-6.

12. Vassilev, L.T., et al., *In vivo Activation of the p53 Pathway by Small-Molecule Antagonists of MDM2.* Science, 2004. **303**(5659): p. 844-8.

13. Shimizu, *Metabolic Regulation of a Bacterial Cell System with Emphasis on Escherichia coli Metabolism.* ISRN Biochemistry, 2013: p. 47.

14. Butcher, E.C., E.L. Berg, and E.J. Kunkel, *Systems Biology in Drug Discovery.* Nature Biotechnology, 2004. **22**(10): p. 1253-9.

15. Bruggeman, F.J. and H.V. Westerhoff, *The Nature of Systems Biology.* Trends in Microbiology, 2007. **15**(1): p. 45-50.

16. Ideker, T., T. Galitski, and L. Hood, *A New Approach to Decoding Life: Systems Biology.* Annual Review of Genomics and Human Genetics, 2001. **2**: p. 343-72.

17. Weckwerth, W., *Metabolomics in Systems Biology.* Annual Review of Plant Biology, 2003. **54**: p. 669-89.

18. Patti, G.J., O. Yanes, and G. Siuzdak, *Innovation: Metabolomics: The Apogee of the Omics Trilogy.* Nature Reviews. Molecular Cell Biology, 2012. **13**(4): p. 263-9.

19. Johnson, C.H. and F.J. Gonzalez, *Challenges and Opportunities of Metabolomics.* Journal of Cellular Physiology, 2012. **227**(8): p. 2975-81.

20. Fiehn, O., *Metabolomics - The Link Between Genotypes and Phenotypes.* Plant Molecular Biology, 2002. **48**(1-2): p. 155-171.

21. Oliver, S.G., et al., *Systematic Functional Analysis of the Yeast Genome.* Trends in Biotechnology, 1998. **16**(9): p. 373-8.

22. Harris, E.D., *Biochemical Facts Behind the Definition and Properties of Metabolites.* FDA.

23. Kanehisa, M. and S. Goto, *KEGG: Kyoto Encyclopedia of Genes and Genomes.* Nucleic Acids Research, 2000. **28**(1): p. 27-30.

24. Caspi, R., et al., *The MetaCyc Database of Metabolic Pathways and Enzymes and the BioCyc Collection of Pathway/Genome Databases.* Nucleic Acids Research, 2012. **40**: p. D742-53.

25. Zhang, G.F., et al., *Metabolomics, Pathway Regulation, and Pathway Discovery.* The Journal of Biological Chemistry, 2011. **286**(27): p. 23631-5.

26. Kell, D.B. and S.G. Oliver, *Here is the Evidence, Now What is the Hypothesis? The complementary Roles of Inductive and Hypothesis-Driven Science in the Post-Genomic Era.* BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology, 2004. **26**(1): p. 99-105.

27. Nicholls, A.W., *Realising the Potential of Metabolomics.* Bioanalysis, 2012. **4**(18): p. 2195-7.

28. Kim, H.K., Y.H. Choi, and R. Verpoorte, *NMR-Based Metabolomic Analysis of Plants.* Nature Protocols, 2010. **5**(3): p. 536-49.

29. Kim, H.K. and R. Verpoorte, *Sample Preparation for Plant Metabolomics.* Phytochemical Analysis : PCA, 2010. **21**(1): p. 4-13.

30. Tolstikov, V.V. and O. Fiehn, *Analysis of Highly Polar Compounds of Plant Origin: Combination of Hydrophilic Interaction Chromatography and Electrospray Ion Trap Mass Spectrometry.* Analytical Biochemistry, 2002. **301**(2): p. 298-307.

31. van der Werf, M.J., et al., *Microbial Metabolomics: Toward a Platform with Full Metabolome Coverage.* Analytical Biochemistry, 2007. **370**(1): p. 17-25.

32. Zhang, B. and R. Powers, *Analysis of Bacterial Biofilms using NMR-Based Metabolomics.* Future Medicinal Chemistry, 2012. **4**(10): p. 1273-306.

33. Olivier, I. and T. Loots du, *A Metabolomics Approach to Characterise and Identify Various Mycobacterium Species.* Journal of Microbiological Methods, 2012. **88**(3): p. 419-26.

34. Villas-Boas, S.G., et al., *Global Metabolite Analysis of Yeast: Evaluation of Sample Preparation Methods.* Yeast, 2005. **22**(14): p. 1155-69.

35. van der Greef, J., P. Stroobant, and R. van der Heijden, *The Role of Analytical Sciences in Medical Systems Biology.* Current Opinion in Chemical Biology, 2004. **8**(5): p. 559-65.

36. Jones, O.A. and V.L. Cheung, *An Introduction to Metabolomics and its Potential Application in Veterinary Science.* Comparative Medicine, 2007. **57**(5): p. 436-42.

37. Sabatine, M.S., et al., *Metabolomic Identification of Novel Biomarkers of Myocardial Ischemia.* Circulation, 2005. **112**(25): p. 3868-75.

38. van Ravenzwaay, B., et al., *The Use of Metabolomics for the Discovery of New Biomarkers of Effect.* Toxicology Letters, 2007. **172**(1-2): p. 21-8.

39. Nicholson, J.K., et al., *Metabonomics: a Platform for Studying Drug Toxicity and Gene Function.* Nature Reviews. Drug Discovery, 2002. **1**(2): p. 153-61.

40. Powers, R., *NMR Metabolomics and Drug Discovery.* Magnetic Resonance in Chemistry : MRC, 2009. **47 Suppl 1**: p. S2-11.

41. Wishart, D.S., *Applications of Metabolomics in Drug Discovery and Development.* Drugs in R&D, 2008. **9**(5): p. 307-22.

42. Beger, R.D., J. Sun, and L.K. Schnackenberg, *Metabolomics Approaches for Discovering Biomarkers of Drug-Induced Hepatotoxicity and Nephrotoxicity.* Toxicology and Applied Pharmacology, 2010. **243**(2): p. 154-66.

43. Zhang, A., et al., *Metabonomics for Discovering Biomarkers of Hepatotoxicity and Nephrotoxicity.* Die Pharmazie, 2012. **67**(2): p. 99-105.

44. Dunn, W.B. and D.I. Ellis, *Metabolomics: Current Analytical Platforms and Methodologies.* Trends in Analytical Chemistry, 2005. **24**(4): p. 285-294.

45. Stoyanova, R. and T.R. Brown, *NMR Spectral Quantitation by Principal Component Analysis. III. A Generalized Procedure for Determination of Lineshape Variations.* Journal of Magnetic Resonance, 2002. **154**(2): p. 163-75.

46. Bylesjo, M., et al., *OPLS Discriminant Analysis: Combining the Strengths of PLS-DA and SIMCA Classification.* Journal of Chemometrics, 2006. **20**(8-10): p. 341-351.

47. Forgue, P., et al., *NMR Metabolic Profiling of Aspergillus Nidulans to Monitor Drug and Protein Activity.* Journal of Proteome Research, 2006. **5**(8): p. 1916-23.

48. Westerhuis, J.A., et al., *Assessment of PLSDA Cross Validation.* Metabolomics, 2008. **4**(1): p. 81-89.

49. Halouska, S., et al., *Predicting the in Vivo Mechanism of Action for Drug Leads Using NMR Metabolomics.* ACS chemical biology, 2011.

50. Wishart, D.S., et al., *HMDB 3.0--The Human Metabolome Database in 2013.* Nucleic acids research, 2013. **41**: p. D801-7.

51. Cui, Q., et al., *Metabolite Identification via the Madison Metabolomics Consortium Database.* Nature Biotechnology, 2008. **26**(2): p. 162-4.

52. Akiyama, K., et al., *PRIMe: A Web Site that Assembles Tools for Metabolomics and Transcriptomics.* In Silico Biology, 2008. **8**(3-4): p. 339-45.

53. Hu, K., W.M. Westler, and J.L. Markley, *Simultaneous Quantification and Identification of Individual Chemicals in Metabolite Mixtures by Two-Dimensional Extrapolated Time-Zero (1)H-(13)C HSQC (HSQC(0))*. Journal of the American Chemical Society, 2011. **133**(6): p. 1662-5.

54. Zhang, B., et al., *NMR Analysis of a Stress Response Metabolic Signaling Network.* Journal of Proteome Research, 2011. **10**(8): p. 3743-54.

55. Halouska, S., et al., *Use of NMR Metabolomics to Analyze the Targets of D-Cycloserine in Mycobacteria: Role of D-Alanine Racemase.* Journal of Proteome Research, 2007. **6**(12): p. 4608-14.

56. Nuxoll, A.S., et al., *CcpA Regulates Arginine Biosynthesis in Staphylococcus aureus Through Repression of Proline Catabolism.* PLoS Pathogens, 2012. **8**(11): p. e1003033.

57. Gebregiworgis, T., et al., *Potential of Urinary Metabolites for Diagnosing Multiple Sclerosis.* ACS Chemical Biology, 2013. **8**(4): p. 684-90.

58. Chaika, N.V., et al., *MUC1 Mucin Stabilizes and Activates Hypoxia-Inducible Factor 1 Alpha to Regulate Metabolism in Pancreatic Cancer.* Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**(34): p. 13787-92.

59. Werth, M.T., et al., *Analysis of Metabolomic PCA Data using Tree Diagrams.* Analytical Biochemistry, 2010. **399**(1): p. 58-63.

60. Worley, B., S. Halouska, and R. Powers, *Utilities for Quantifying Separation in PCA/PLS-DA Scores Plots.* Analytical Biochemistry, 2013. **433**(2): p. 102-4.

61. Vuong, C. and M. Otto, *Staphylococcus epidermidis Infections.* Microbes and infection / Institut Pasteur, 2002. **4**(4): p. 481-9.

62. Ray, G.T., J.A. Suaya, and R. Baxter, *Trends and Characteristics of Culture-Confirmed Staphylococcus aureus Infections in a Large U.S. Integrated Health Care Organization.* Journal of Clinical Microbiology, 2012. **50**(6): p. 1950-7.

63. Otto, M., *Staphylococcal Infections: Mechanisms of Biofilm Maturation and Detachment as Critical Determinants of Pathogenicity.* Annual Review of Medicine, 2013. **64**: p. 175-88.

64. Mack, D., et al., *The Intercellular Adhesin Involved in Biofilm Accumulation of Staphylococcus epidermidis is a Linear Beta-1,6-Linked Glucosaminoglycan: Purification and Structural Analysis.* Journal of Bacteriology, 1996. **178**(1): p. 175-83.

65. Gotz, F., *Staphylococcus and Biofilms.* Molecular Microbiology, 2002. **43**(6): p. 1367-78.

66. Park, A., et al., *Effect of Shear Stress on the Formation of Bacterial Biofilm in a Microfluidic Channel.* Biochip Journal, 2011. **5**(3): p. 236-241.

67. Stewart, P.S. and J.W. Costerton, *Antibiotic Resistance of Bacteria in Biofilms.* Lancet, 2001. **358**(9276): p. 135-8.

68. Kaplan, J.B., et al., *Enzymatic Detachment of Staphylococcus epidermidis Biofilms.* Antimicrobial Agents and Chemotherapy, 2004. **48**(7): p. 2633-6.

69. Seidl, K., M. Bischoff, and B. Berger-Bachi, *CcpA Mediates the Catabolite Repression of tst in Staphylococcus aureus.* Infection and Immunity, 2008. **76**(11): p. 5093-9.

70. Sadykov, M.R., et al., *CcpA Coordinates Central Metabolism and Biofilm Formation in Staphylococcus epidermidis.* Microbiology, 2011. **157**(Pt 12): p. 3458-68.

71. Koch, A., ed. *The Bacteria: Their Origin, Structure, Function and Antibiosis*. 2006, Springer.

72. Mitnick, C.D., et al., *Comprehensive Treatment of Extensively Drug-Resistant Tuberculosis.* The New England Journal of Medicine, 2008. **359**(6): p. 563-74.

73. Cunha, B.A., *Antibiotic side effects.* The Medical Clinics of North America, 2001. **85**(1): p. 149-85.

74. Torun, T., et al., *Side Effects Associated with the Treatment of Multidrug-Resistant Tuberculosis.* The International Journal of Tuberculosis and Lung Disease : The Official Journal of the International Union against Tuberculosis and Lung Disease, 2005. **9**(12): p. 1373-7.

75. Feng, Z. and R.G. Barletta, *Roles of Mycobacterium smegmatis D-Alanine:D-Alanine Ligase and D-Alanine Racemase in the Mechanisms of Action of and Resistance to the Peptidoglycan Inhibitor D-Cycloserine.* Antimicrobial Agents and Chemotherapy, 2003. **47**(1): p. 283-91.

76. Caceres, N.E., et al., *Overexpression of the D-alanine Racemase Gene Confers Resistance to D-Cycloserine in Mycobacterium smegmatis.* Journal of Bacteriology, 1997. **179**(16): p. 5046-55.

77. Zygmunt, W.A., *Antagonism of D-Cycloserine Inhibition of Mycobacterial Growth by D-Alanine.* Journal of Bacteriology, 1963. **85**: p. 1217-20.

78. Chacon, O., et al., *Mycobacterium smegmatis D-Alanine Racemase Mutants Are Not Dependent on D-Alanine for Growth.* Antimicrobial Agents and Chemotherapy, 2002. **46**(1): p. 47-54.

79. Milligan, D.L., et al., *The Alanine Racemase of Mycobacterium smegmatis is Essential for Growth in the Absence of D-Alanine.* Journal of Bacteriology, 2007. **189**(22): p. 8381-6.

CHAPTER 2

TOOLS FOR INTERPRETTING NMR MULTIVARIATE DATA

2.1 Introduction

The use of pattern recognition to analyze complex data sets has been one of the most important developments in the application of NMR to metabolomics. NMR data are generally large and complex, consisting of thousands of signals corresponding to numerous metabolites in a biological sample. Thus, the aim of multivariate statistical techniques is to maximize the interpretation of NMR datasets by identifying and quantifying all the metabolites that are perturbed by genetic alterations, disease processes, or drug treatment [1]. However, the numerous metabolite signals tend to result in a congested NMR spectrum, which hinders manual interpretation of the data.

Pattern recognition tools, such as principal component analysis (PCA) and orthogonal projection of latent structures discriminant analysis (OPLS-DA), are routinely used to analyze NMR datasets collected on metabolomes extracted from various cell cultures. The resulting PCA or OPLS-DA scores plots provide a means to identify biological relationships between the different cell cultures based on global differences or similarities between the cellular metabolomes [2, 3]. In general, pattern recognition tools are used to identify the spectral features that significantly contribute to a differentiation between the classes in an attempt to discover biomarkers or to construct a model to determine class membership [4, 5]. PCA is an unsupervised technique, which means that class designations are not used to generate a model [6]. Instead the model and the class

discrimination is based on identifying the largest variations in the data [6]. However, PCA is only applicable if variations in the experimental dataset are a consequence of differences in an underling biological process. In effect, differences in the NMR spectra must result from biologically relevant changes in the cellular metabolomes.

Alternatively, a supervised technique, such as OPLS-DA, requires class designations in order to generate a model. Similar to unsupervised techniques, supervised methods only generate a *validated* model if a difference in the metabolome and associated NMR spectra actually exist. Unlike unsupervised techniques, supervised methods can also generate the appearance of class separation even for completely random data [7]. Thus, it is essential that the model generated by a supervised technique is properly validated [7]. Despite these added challenges, the advantage of supervised techniques is the ease of identifying the key spectral features (metabolites) that define the class separation.

In this chapter, I present an overview, and describe the general application of PCA and OPLS-DA for the analysis of NMR metabolomics data as a tool for systems biology. I will also highlight the importance of generating statistically meaningful OPLS-DA models using validations protocols. In addition, I will briefly describe the statistical tools PCAtoTree and PCA/PLS-DA utilities (http://bionmr.unl.edu/pca-utils.php) that have been developed by Prof. Mark Werth and Bradley Worley, respectively, with the assistance of my NMR metabolomics datasets, to quantify the statistical significance of class separation in PCA and OPLS-DA scores plots. Finally, I

will describe the application of OPLS-DA S-plots to identify the significant metabolite changes that are responsible for class separation.

2.2 Principal Component Analysis

PCA is a well-established unsupervised statistical method for interpreting large multivariable datasets. PCA results in the reduction of an NMR spectrum into a single point in a Cartesian coordinate system in a K-dimensional space, where each dimension represents a chemical shift bin from the NMR spectra (Figure 2.1) [8]. The NMR data matrix X consists of *n* samples and *m* variables (chemical shift bins) which can be represented as an ensemble of *n* points in an *m* dimensional space. PCA transforms the NMR data matrix X by a partial linear least squares fit to a principal component vector ($\overrightarrow{PC_1}$) corresponding to the largest variation in the dataset. The data is then fit to a second orthogonal vector ($\overrightarrow{PC_2}$) corresponding to the second largest variation in the dataset. Each successive principal component vector describes a decreasing amount of the variability in the NMR data matrix X. A "weights" or "loadings" matrix (P) maps the NMR data matrix X onto these principal component vectors. Thus, the new transformed data will have a new set of axes which relates to each principal component through a scores matrix (T). In essence, PCA performs an orthogonal linear transformation of a multivariate dataset into the latent (PC) coordinate space while maintaining most of the variability of the original data (Equation 2.1)[9].

$$T = XP \tag{2.1}$$

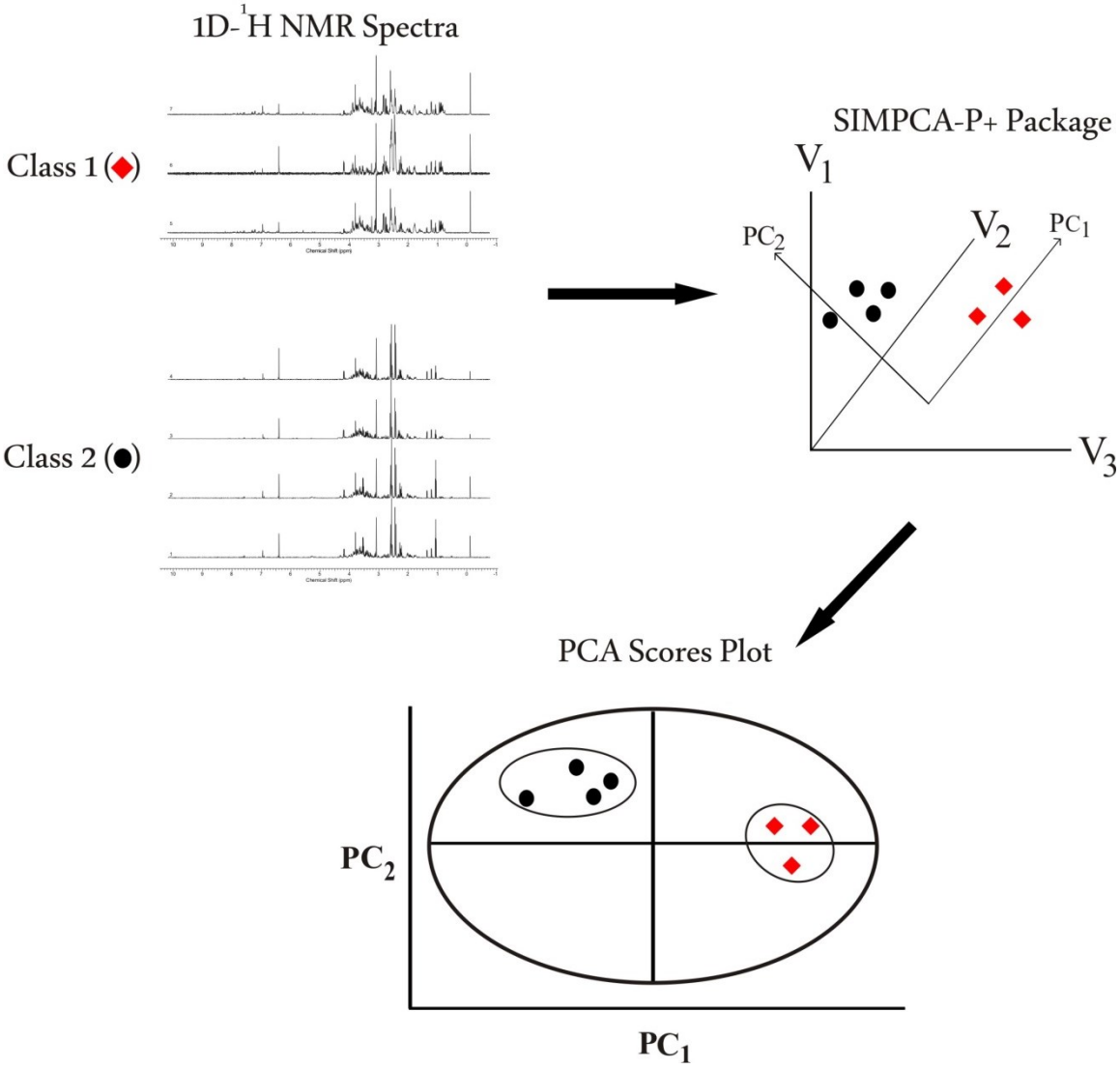


Figure 2.1. Diagram of 1D ^{1}H NMR data being reduced into a PCA scores plot. Each spectrum is reduced to a single point along the k-dimensional variable space. Each dimensional variable represents a chemical shift bin. PC1 account for the largest variation, and PC2 is orthogonal to PC1 and describes the second largest variation. Reprinted with permission [8] © Future Medicinal Chemistry, 2012.

PCA can also be generally described as a method that decomposes the NMR data matrix (X) into two smaller matrices corresponding to the scores matrix (T) and a transposed loading matrix (P^T) (Equation 2.2) [10, 11],

$$X = TP^T + E \tag{2.2}$$

The scores matrix captures the systematic variation in the dataset with respect to the samples and variables. The loadings matrix captures the relative contribution of each chemical shift bin to this variance. The E matrix contains the residuals, the data in X that is not explained by the PCA model. The principal components (PC) are a series of partial linear least squares fits of the original set of samples in X that captures the maximum amount of variance in the dataset. Therefore equation 2.2 can be expanded as shown below [10, 11],

$$X = t_1p_1^T + t_2p_2^T + t_3p_3^T + \cdots + t_Kp_K^T + E \tag{2.3}$$

where t_K and p_K^T represent the scores vector and loadings vector for each successive principal component. The first principal component ($t_1p_1^T$) explains the largest amount of variance in the data. The second component ($t_2p_2^T$) is orthogonal to the first component and explains the second largest amount of variance in the data. Each subsequent PC_K will capture a decreasing amount of variance.

2.3 Orthogonal Projection of Latent Structure

PCA can be advantageous for detecting outliers, finding patterns or trends, or the separation of groups that are not similar [12, 13]. This means PCA can be performed with little to no knowledge about the samples being studied. However, PCA based

approaches are limited, because the unsupervised approach results in a separation based only on the maximum variations [3]. Thus, the principal components may not necessarily reflect differences between classes [14]. On the other hand, OPLS-DA is a supervised technique where the results in an OPLS-DA scores plot is driven by the class assignments [3, 15]. This means that the model is rotated based on the class assignments instead of the maximum variations in the data set (Figure 2.2). More specifically, the first (predictive) component of the OPLS-DA scores spectra contains information that distinguishes between class separations, while the second (orthogonal) component contains information relative to within class membership**.**

OPLS-DA is a supervised algorithm that uses an orthogonal signal correction (OSC) filter to remove any variation in data matrix X that is not correlated with a response matrix Y [3, 16]. The data matrix Y contains discrete classifications, such as wild-type, mutant, drugs, or disease. As an illustrative example, the wild-type cultures can be assigned a value of 0, whereas the mutant, drugs, or disease can be designated a value of 1. However, the class designation can be used to compare any class combinations such as comparison between different mutants or drug treatments. Overall, OPLS-DA uses information of data matrix Y to decompose data matrix X into blocks of structured variations that are correlated to and orthogonal to Y, respectively using the following equation [3, 11],

$$X = T_p P_p^T + T_o P_o^T + E \tag{2.4}$$

where T_p and P_p^T are the predictive scores and loadings that are correlated with Y. T_o and P_o^T are the scores and loadings, respectively, that are orthogonal to Y. E is the residuals

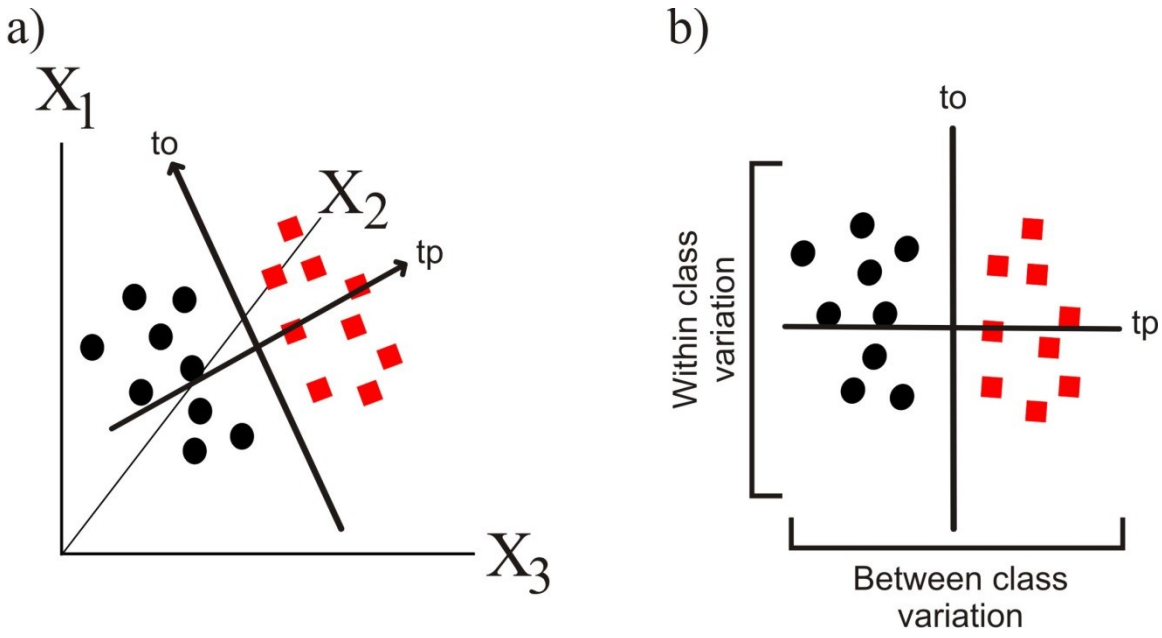


Figure 2.2. Diagram of 1D ^{1}H NMR data being reduced into an OPLSA-DA scores plot. The OPLS-DA scores plot is rotated based on the classification assignment where the predictive component (tp) represent between class variations and the orthogonal component (to) represents within class variation.

or variations in X that are not explained in the model. Also in the above expression, the $T_pP_p^T$ block (predictive components) represents variations between class memberships. $T_oP_o^T$ (orthogonal components) represents the within class variations.

2.4 Interpretation of PCA and OPLS-DA models

The use of PCA in NMR metabolomics is an effective strategy for obtaining useful information about efficacy, toxicity, and selectivity of chemical leads in the drug discovery process [17-19]. One simple approach to determine the selectivity and toxicity of a drug is through the comparison of the metabolome from a wild-type strain, a mutant strain, and wild-type and mutant strains treated with a chemical lead [18]. The mutant strain must consist of an inactivated or diminished activity for the protein hypothesized to be the target of the chemical lead. Also, the metabolome from the mutant strain must show good separation from the metabolome of the wild type cells in a PCA scores plot. The mutant and wild-type cultures are then treated with the chemical lead, resulting in a total of four groups (wild-type cells, drug-treated wild-type cells, mutant cells, and drug-treated mutant cells). Depending on the drugs *in vivo* activity, four outcomes in the PCA scores plot are possible (Figure 2.3). If the PCA scores plot results in two separate clusters where the metabolomes from the drug-treated cells do not separate from the wild-type or mutant metabolomes, then the chemical lead is said to be inactive (Figure 2.3a). Conversely, the chemical lead is predicted to be active and selective if the metabolomes from both sets of drug-treated cultures are clustered together with the mutant metabolome in the PCA scores plot (Figure 2.3b). This means genetically or

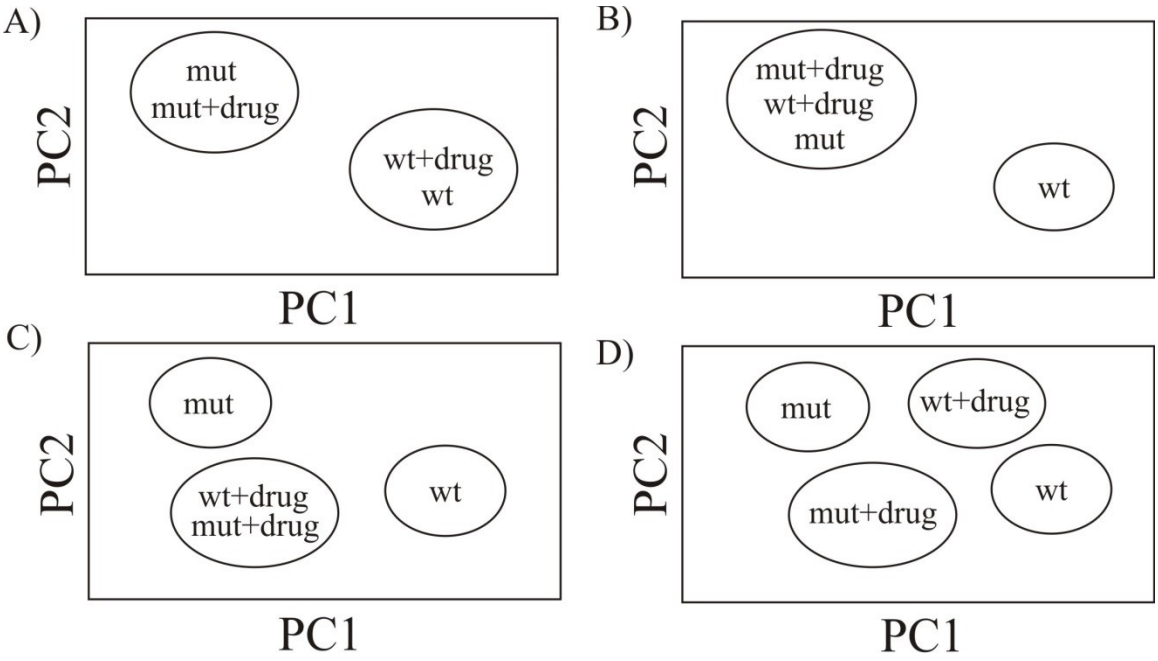


Figure 2.3. Illustration demonstrating the hypothetical PCA scores plot for the following scenarios a) inactive compound, b) active and selective inhibitor, c) active, nonselective inhibition of target and secondary protein, and d) active, nonselective preferential inhibition of secondary protein. Reprinted with permission [18] © Journal of Proteome Research, 2006.

chemically inhibiting the same protein target leads to an essentially identical change in the metabolome. If the PCA scores plots yields three sets of clusters where the metabolomes from both drug-treated cultures are clustered together but are separated from the metabolomes from both the mutant and wild-type cells (Figure 2.3c), the chemical lead is predicted to inhibit two or more protein targets. The drug is active, but not selective. Lastly, if four distinct clustering patterns are formed in the PCA scores plot, then the chemical lead is inhibiting a protein not related to the original, hypothesized protein target (Figure 2.3d).

As an example, Figure 2.4 demonstrates the use of NMR metabolomics methodology to monitor the *in vivo* activity of 8-azaxanthine [18]. 8-Azaxanthine is a well-known inhibitor of urate oxidase in *Aspergillus nidulans* [20, 21]. A comparison was made between the metabolome from *uaZ14* mutant with an inactive urate oxidase, and the wild-type metabolome [22]. Both strains were also treated with 8-azaxanthine, where the resulting PCA scores plot shows two distinct clustering patterns. Both metabolomes from wild-type and *uaZ14* mutant mycelia treated with 8-azaxanthine cluster together with the metabolome from the untreated *uaZ14* mutant mycelia. The metabolome from the untreated wild-type mycelia form a separate cluster in the PCA scores plots. This demonstrates that 8-azaxanthine is a selective inhibitor targeting urate oxidase. Importantly, the PCA can be related back to phenotype, where the wild type mycelia showed hyphal growth that was lacking in *uaZ14* mutant mycelia or 8-azaxanthine treated mycelia [18].

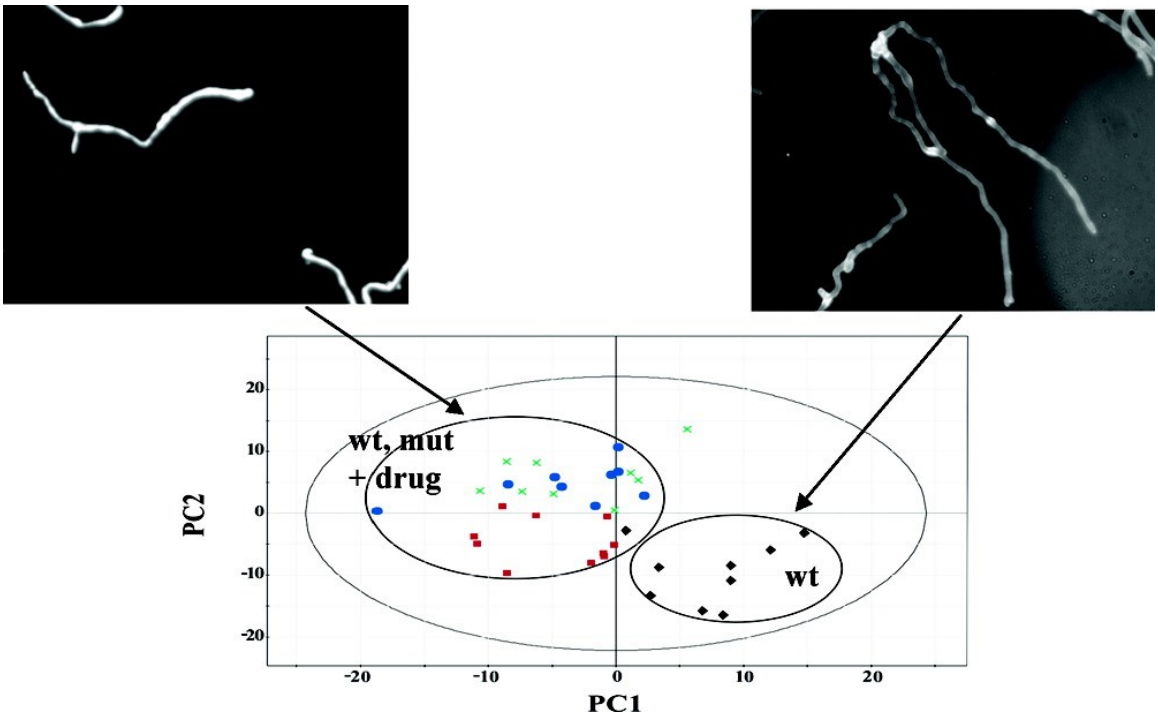


Figure 2.4. PCA scores plot comparing *A. nidulans uaz14* mutant (x), wild-type with 8-azaxanthine (■), *uaZ14 mutant* with 8-azaxanthine (●), and wild-type cells (⧫). The scores plot is then used to compare lectin stains of untreated A. nidulans fungus (left) vs. the fungus treated with 8-azaxanthine (right). The lectin stains, which bind to growing chitin cell was, indicates the amount of hyphal growth that occurred after the stain. Reprinted with permission [18] © Journal of Proteome Research, 2006.

A common problem with the evaluation of a PCA model is the identification of classes or groups, such as the separation between a control group and a group being observed to have a specific disease or origin. Replicate metabolomics samples tend to exhibit a significant variation between each sample due to differences in sample growth, sample handling, data collection, and the inherent variability in nature. Although it is important to minimize these variations, it is impractical to maintain perfect consistency. Instead, OPLS-DA is able to improve class separation, because variations not correlated with a class separation are simply removed. As an illustration, Figure 2.5 displays a comparison between PCA and OPLS-DA scores plot used to discriminate between strains of *Scutellana baicalensis* bacteria isolated from different locales [23]. The PCA scores plot showed some class differences, but there is noticeable overlap due to within group variations. This makes classification a challenge because it is not clear if the two groups are statistically different. The separation of the NMR metabolomics data into predictive and orthogonal components in OPLS-DA scores plot allows for the identification of class discrimination (Figure 2.5b). This would enable verification of the statistical model by predicting the origins of other *S. baicalensis* isolates. However, extreme care must be taken because OPLS-DA has a tendency to over-fit the data [24]. As a result, OPLS-DA can generate scores plots with an apparent class separation even for randomly generated data [7]. Thus, the validation of every OPLS-DA model is essential.

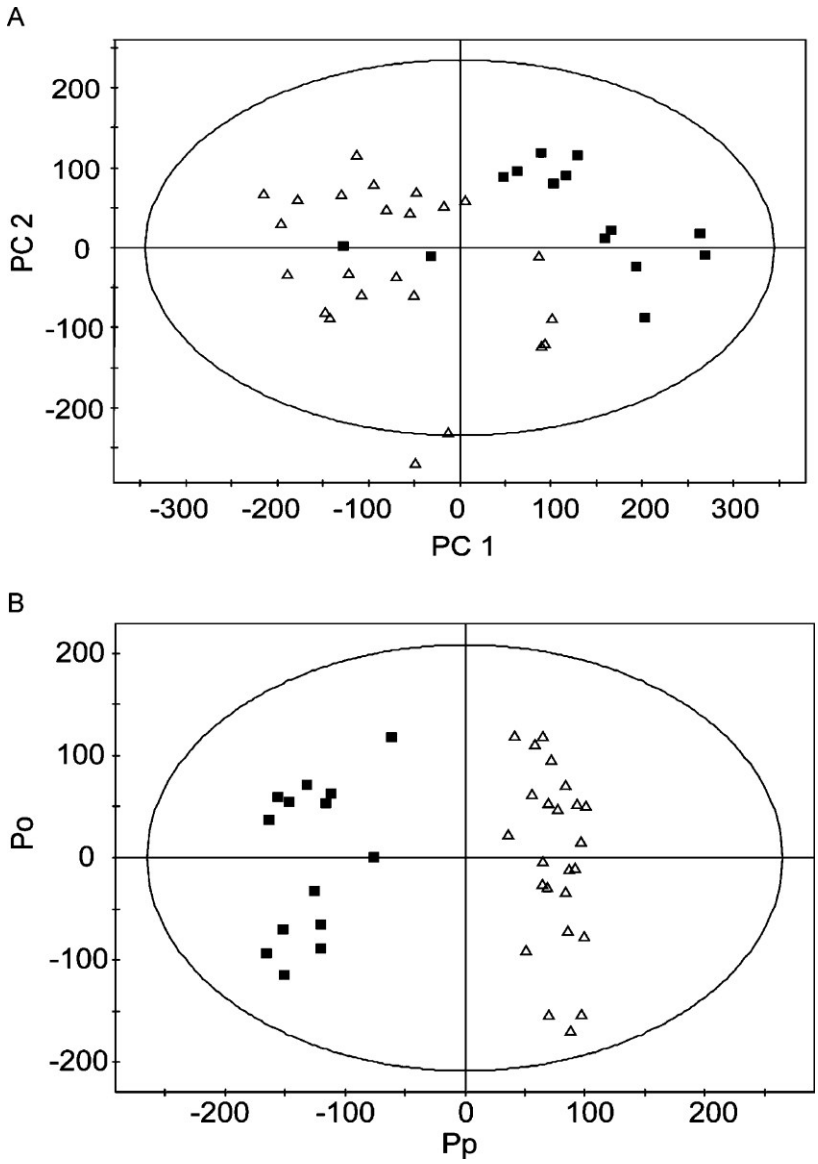


Figure 2.5. a) PCA and b) OPLS-DA scores plot comparing the metabolic fingerprints of Korean (Δ) and Chinese *S. baicalensis* groups (■). Reprinted with permission [23] © Journal of Agricultural and Food Chemistry, 2008.

2.5 Validation of PCA and OPLS-DA models

One of the main objectives for PCA and OPLS-DA is to select a model that best represents the data. Since the model is heavily simplified for the ease of visual inspection, the model might give rise to an unclear or false positive result [25]. Therefore, statistical analysis is needed in order to establish a valid and meaningful model. For both PCA and OPLS-DA, a quantitative measure of the goodness of fit is given by the contribution score or the variance that is being explained by each component and is represented as R^2. The goodness of fit is calculated as follows [26, 27],

$$R^2 = 1 - \frac{RSS}{SSX_{TOT}} \tag{2.5}$$

where *RSS* is residual sum of squares and SSX_{TOT} is the total variation in the data. A R^2 of 1.0 represents a complete explanation of the model while a R^2 of 0.0 represents no explanation of the model. Typically, a good model is represented by high contribution scores for a few components, and lower contribution scores are usually attributed to noisy data. R^2 is a good way of quantifying variables being explained for each component. For example, some variables can be explained well by either the first, second, or third component, while some variables may only be explained by multiple component. This allows for a qualitative determination of the variables important for each component.

The goodness of prediction (Q^2) is an estimate of the predictive ability of the model based on cross validation [27]. Cross validation is a predictive strategy of evaluating and comparing the models generated by dividing data into two parts, a training set and validation set [28]. More specifically a certain number of observations are removed from the original dataset and moved into a validation set while the remaining

observations are left in the training set. During the process, a number of successive rounds are performed where the training set and validation set must cross-over so that each point has a chance to be evaluated. There are multiple strategies for cross validation including resubstitution validation, hold-out validation, k-fold cross validation, and leave-one-out cross-validation, and repeated k-fold fold cross-validation [29]. The Simca P+ package uses a type of leave-n-out cross-validation, where $1/7^{th}$ of the data is left out and then examined [27]. The quality of the model can then be estimated using equation 2.6 [26, 27]:

$$Q^2 = 1 - \frac{PRESS}{SSX_{TOT}} \tag{2.6}$$

where *PRESS* is the predictive residual sum of squares determined using the cross-validation models. Overall Q^2 renders a measure of the final model's predictive capabilities. If the Q^2 of each successive component is roughly equal to its R^2, then the components are said to be useful and predictive in the model. A Q^2 score that is "small" compared to the R^2 indicates the component is not predictable and is comparable to fitting a line through randomly generated data or noise.

It is easy to get a relatively high R^2-value in a model by simply adding more components until the value approaches one, since the R^2 of the model is a summation of the R^2 values for each component. However, this leads to an over-fitting of the model. Therefore, it raises the general question: how many components are relevant? In contrast, the Q^2 value for a model will reach a plateau and then begin to decrease for each successive additional component (Figure 2.6) [30]. The decrease in Q^2 indicates that the

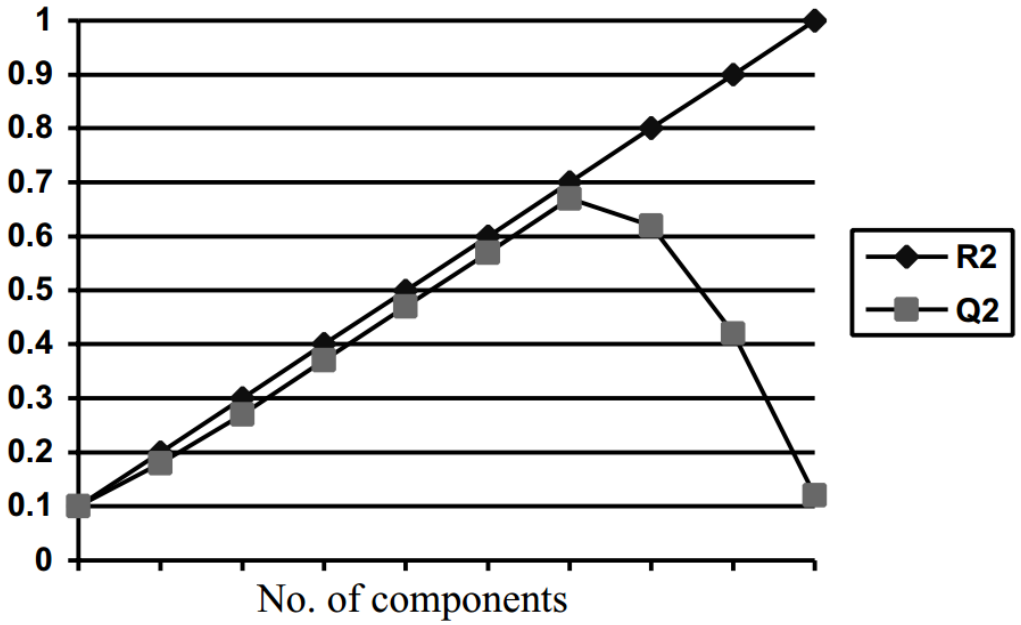


Figure 2.6. The relation between R^2 (goodness of fit) and Q^2 (goodness of prediction) when adding new components in a model. Reprinted with permission [30] © Scandinavian Journal of Psychology, 2001.

additional components are not predictive and are not used [30]. Hence, it is important to determine the optimal number of components by the relationship of the R^2 and Q^2 values.

Although the R^2 and Q^2 values display the overall quality of the model, they do not validate if the class prediction between different groups used in an OPLS-DA model are statistically different. OPLS-DA has a tendency to over-fit the data, where randomly generated groups can generate separate clusters in a scores plot [7]. A permutation test could be used to examine the degree of over-fitting [7]. The method scrambles the classification matrix Y while the original data matrix X remains intact. A new model is then calculated based on the permuted Y matrix, and the permuted model is examined based on the R^2 and Q^2 values. This procedure is repeated multiple times (generally more than 100 times), creating models with different correlations to the original model.

Normally the R^2 and Q^2 values for the permuted models are lower than the original model as a consequence of the reduce correlation to the original Y matrix. The degree of reduction of these values can be used to validate the significance of the group classification. For example, Figure 2.7 is the permutation plot of the OPLS-DA scores plot shown in Figure 2.5b [23]. When examining the R^2 and Q^2 values, the models have been permuted so much that the values can be extrapolated to a zero correlation. The R^2 value at zero correlation in the plot is lower, but does not approach zero, meaning that it is mathematically possible to reproduce any possible combinations of Y with an R^2 of 0.4. The Q^2 value, on the other hand, ended up with a negative value, meaning the predictive quality is extremely poor at a zero correlation. Thus, the original model is well

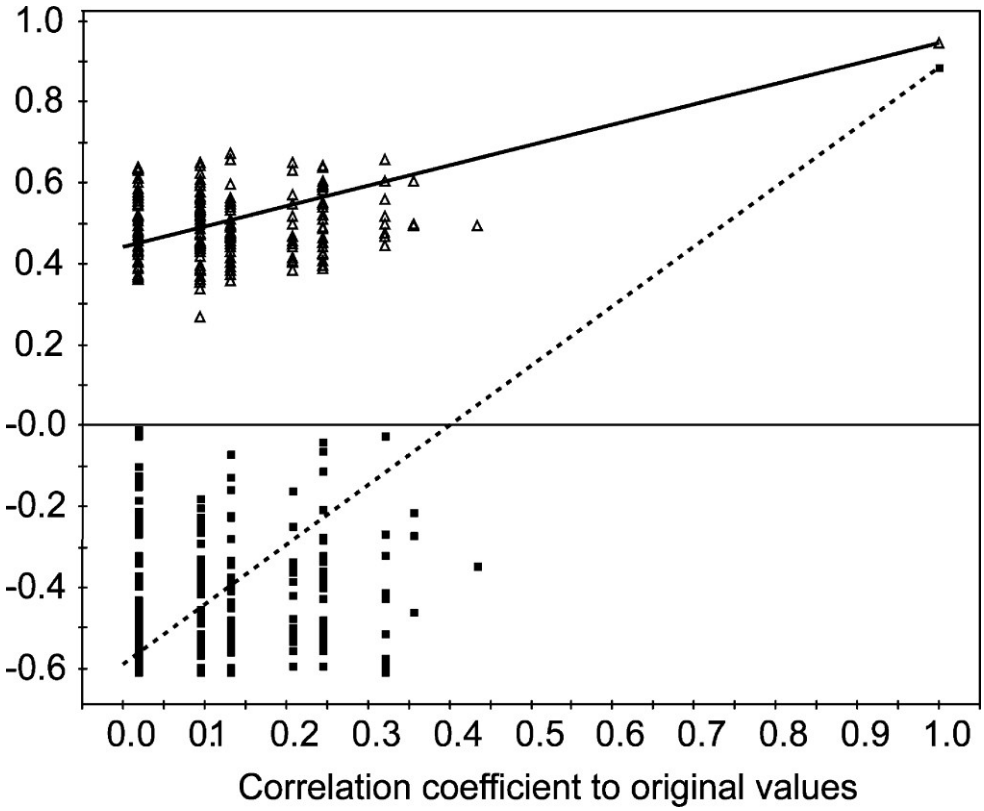


Figure 2.7. Permutation plot describing the results for 200 permutations. The solid line represents the regression line for R^2. The dashed line represents the regression line for Q^2. Reprinted with permission [23] © Journal of Agricultural and Food Chemistry, 2008.

fit and the two groups are significantly different due to the loss of predictive power when the classification is randomly assigned.

Cross Validated-Analysis of Variance (CV-ANOVA) is another method to test the significance between groups [31]. It is based on the cross-validated predictive residuals and predictive scores values, where the model is used to compare the variances of the residuals based on the F-test. The CV-ANOVA provides a wealth of information about the model including the corresponding mean squares, sum of squares, F-values, and p-values. A common practice is to interpret a p-value lower than 0.05 to designated statistically different groups.

2.6 Identifying metabolite perturbations using S-Plots

One of the common goals in metabolomics is to pinpoint metabolite perturbations that are related to drug toxicity, diseases, genetic modifications, or environmental variations. However, it is extremely difficult to summarize and interpret them by visually inspecting complex 1D-NMR data. The S-plot provides an easier way to interpret the data by combining the covariance and correlation loadings resulting from the predictive component in an OPLS-DA scores plot. The covariance loadings correspond to the magnitude or contribution to the scores. The correlation loadings represent the consistency of the model variables. The two vectors can be calculated as shown below [15]:

$$Cov(t, X_k) = \frac{t^T X_k}{N-1} \tag{2.7}$$

$$Corr(t, X_k) = \frac{Cov(t, X_k)}{s_t s_{X_k}} \tag{2.8}$$

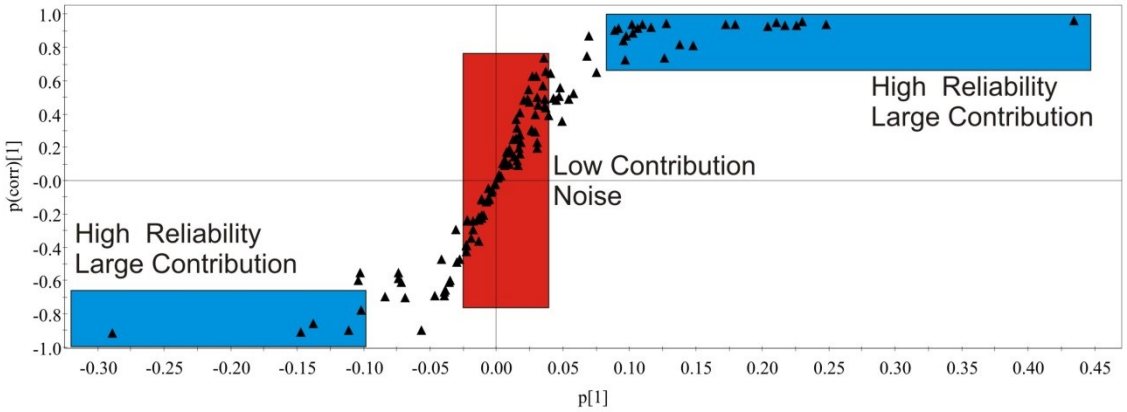


Figure 2.8. Diagram of an S-plot for a given OPLS-DA scores plot. Each point in the S-plot represents a chemical shift bin in a 1D NMR data set. The blue box represents points of interest as they highly contribute to class separation. The red box represents data contributing very little to class separation.

where *t* is the scores vector in the OPLS-DA model, *k* is the centered variable in data matrix X, and *s* is the estimated standard deviations. These two vectors are plotted as a scatter plot and usually form an "S" shape (Figure 2.8). The selection of potentially biochemical interesting compounds is dependent on the location of the point on the S-plot, where each point represents a chemical shift bin, which in turn, can specify a specific or set of metabolites. Points that show high correlation and covariance (blue box) are of interest as they are responsible for class separation. Points that show low correlation are unreliable. Points showing low covariance contribute little to no class separation (red box). One issue with the S-plot is that the magnitude of the contribution is relative, so there is no defined absolute value that designates a significant contribution to class separation. As a conservative approach, points at the extreme regions along the covariance axis are considered points of interest. As an example, in Figure 2.8 points with a covariance greater than 0.1 or less than -0.1 would be of interest. However, identifying points within the high risk region (high correlation, low covariance) can be a daunting task as these metabolites may or may not be important. Therefore, the use of loadings plots from PCA and quantitative analysis of ^{13}C-metabolites using 2D ^{1}H-^{13}C HSQC's is a good complement to identifying important metabolites.

2.7 PCAtoTree and PCA Utilities

PCA and OPLS-DA are able to decompose high-throughput data into qualitative visual representation, showing clustering patterns of biological samples into either similar or different groups. In some cases, the clustering patterns in a scores plot are distinct and

well separated. In other cases, the scores plot clustering patterns are not very well defined. Also, as the number of different groups increases (*i.e.,* different drug treatments or mutations) the scores plot becomes more complex and it becomes difficult to determine if the groups are significantly different. The simple visual inspection of a PCA and OPLS-DA scores plot does not imply the statistical significance of group separation. In turn, a visual inspection of a scores plot does not always provide a reliable interpretation of the biological significance of the group separation. OPLS-DA model validation methods, such as the permutation plot or CV-ANOVA, can only determine if two groups are significantly different. Therefore, PCAtoTree was designed to analyze clusters in a scores plot by generating tree diagrams with corresponding bootstrap numbers to provide a statistical significance to each group separation [32].

PCAtoTree [32] is an awk-based script written by Prof. Mark Werth that uses the data generated by the PCA scores plot and determines the average distance and standard deviation for each assigned group. The program uses a bootstrapping method that generates 100 distance matrices based on the average position of each cluster. The Phylip software package (http://evolution.genetics.washington.edu) utilizes the 100 distance matrices to generate a consensus tree diagram with the corresponding bootstrap values for each node. Bootstrap values below 50% indicate the separation between clusters is insignificant. Figure 2.9 is an example of a 2D PCA scores plot and associated metabolomics tree diagram, showing the relative similarities between the metabolomes from different *Mycobacterium smegmatis* mutants and wild type cells in the presence and absence of D-cycloserine treatment.

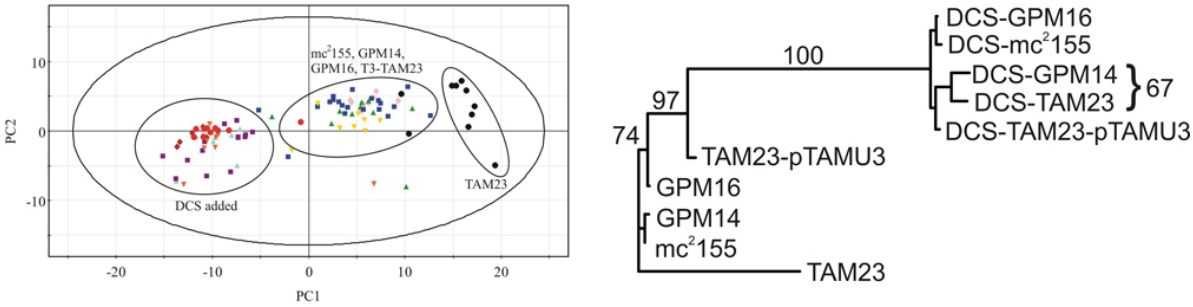


Figure 2.9. a) PCA scores plot comparing mc^{2}155(■), TAM23 (●), GPM14 (♦), GPM 16 (▲) TAM23 pTAMU3 (▽), mc^{2}155 with DCS (■), TAM23 with DCS (●), GPM14 with DCS (♦), GPM16 with DCS (▲) and TAM23pTAMu3 with DCS (▼). Tree diagram corresponding to the PCA scores plot. Bootstrapping numbers less than 50 are not displayed indicating similar clustering. Reprinted with permission [32] © Analytical Biochemistry, 2010.

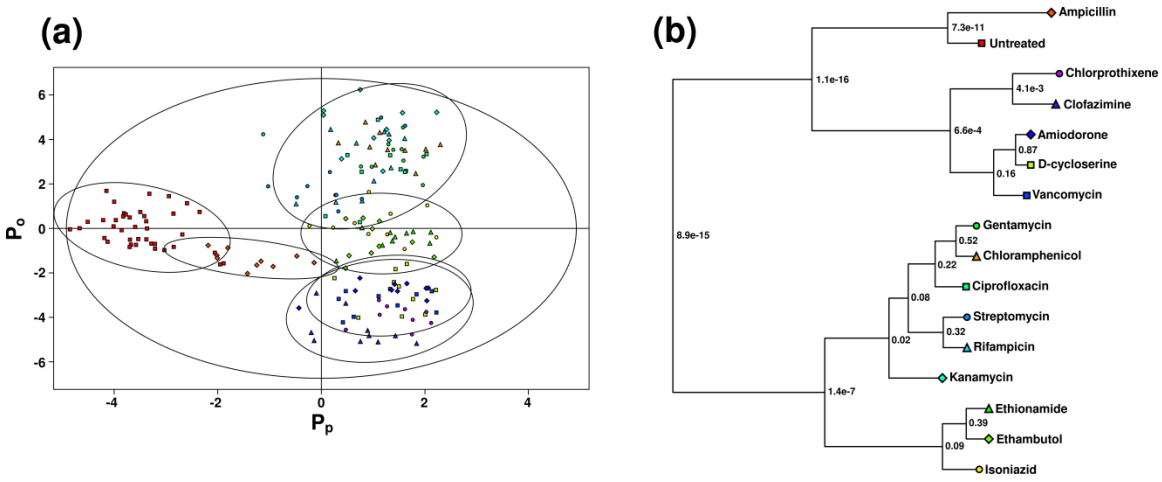


Figure 2.10. (a) 2D OPLS-DA scores plot illustrating 95% confidence ellipses for data having one predictive and one orthogonal PLS component. The symbol shape and color of each point correspond to the groups in (b). Discrimination in the first component is between wild-type and antibiotic-treated *Mycobacterium smegmatis*, and separations along the second component indicate metabolic differences between various antibiotic treatments. The antibiotics cluster together based on a shared biological target (cell wall synthesis, mycolic acid biosynthesis, or transcription, translation and DNA supercoiling). Three compounds of unknown in vivo activity were shown to cluster together with inhibitors of cell wall synthesis, implying a potential biological target. Interestingly, the *M. smegmatis* strain is resistant to ampicillin resulting in the ampicillin-treated cells clustering closer to untreated cells. The ellipses define the statistical significance of class separation and provide an illustration where two groups actually belong to the same biological classification. (b) Dendrogram generated from scores in (a) using Mahalanobis distances, with p values for the null hypothesis reported at each branch. Reprinted with permission [33] © Analytical Biochemistry, 2013.

PCAtoTree has been recently updated to the PCA/PLS-DA utilities written by Bradley Worley [33]. In addition to using the bootstrapping, the program now uses a Mahalanobis distance matrix to generate a T^2 Hotelling and F statistical values. Therefore, the statistical separations between clusters in a scores plot can now be quantified using a p-value. The PCA/PLS-DA utilities can now apply 95% confidence ellipses or ellipsoids around each cluster in a scores plot. Ellipses were previously hand drawn to show visual representations for each cluster, but did not have any statistical meaning and may lead to inaccurate interpretation. Figure 2.10 is an example of the 2D OPLS-DA scores plot illustrating the 95% confidence ellipses defining the different mechanism of actions of 15 different drugs. The associated tree diagrams quantify the clustering pattern, where p-values of less than 0.05 identify statistically significant nodes or group separations.

2.8 Conclusion

Pattern recognition tools such as PCA and OPLS-DA are valuable for interpreting large data sets. This is important in systems biology as the metabolome for each organism is very complex and quite dynamic. PCA and OPLS-DA are able to take advantage of the complexity of the data and reduce them down to simple plots that are easy to examine. Nevertheless, the number of application is small as metabolomics is a relatively new field. However, the potential of pattern recognition tools have yet to reach its full potential as new strategies continue to be developed. The advantage of these tools is its relative simplicity; however, it can be easy to manipulate these plots to gain modest

separation, which may provide incorrect results about perturbation of the system as a whole. For example, relating systematic perturbation to only a few peaks and removing all biologically relevant data should be prohibited as this does not relate to the whole metabolomics system. Hence, much of the information is lost. Especially since metabolic changes of one network may constitute passive alterations of other pathways to maintain metabolic homeostasis [34]. Instead loading plots and S-plots should be used to identify these metabolite perturbations. More importantly, every model should be validated to give a statistical meaningful representation of the data. There are tools such as CV-ANOVA, permutation plots, PCAtoTree, and PCA/PLS-DA utilities that are able to statistically validate cluster separations in a scores plot to yield a reliable interpretation of the biological significance of the metabolome changes.

2.9 References

1. Nicholson, J.K., J.C. Lindon, and E. Holmes, *'Metabonomics': Understanding the Metabolic Responses of Living Systems to Pathophysiological Stimuli via Multivariate Statistical Analysis of Biological NMR Spectroscopic Data.* Xenobiotica; The Fate of Foreign Compounds in Biological Systems, 1999. **29**(11): p. 1181-9.

2. Holmes, E. and H. Antti, *Chemometric Contributions to the Evolution of Metabonomics: Mathematical Solutions to Characterising and Interpreting Complex Biological NMR spectra.* The Analyst, 2002. **127**(12): p. 1549-57.

3. Bylesjo, M., et al., *OPLS Discriminant Analysis: Combining the Strengths of PLS-DA and SIMCA Classification.* Journal of Chemometrics, 2006. **20**(8-10): p. 341-351.

4. Goodacre, R., et al., *Metabolomics by Numbers: Acquiring and Understanding Global Metabolite Data.* Trends in Biotechnology, 2004. **22**(5): p. 245-52.

5. Coen, M., et al., *NMR-Based Metabolic Profiling and Metabonomic Approaches to Problems in Molecular Toxicology.* Chemical Research in Toxicology, 2008. **21**(1): p. 9-27.

6. Ivosev, G., L. Burton, and R. Bonner, *Dimensionality Reduction and Visualization in Principal Component Analysis.* Analytical Chemistry, 2008. **80**(13): p. 4933-44.

7. Westerhuis, J.A., et al., *Assessment of PLSDA Cross Validation.* Metabolomics : Official Journal of the Metabolomic Society, 2008. **4**(1): p. 81-89.

8. Zhang, B. and R. Powers, *Analysis of Bacterial Biofilms using NMR-Based Metabolomics.* Future Medicinal Chemistry, 2012. **4**(10): p. 1273-306.

9. Abdi, H., Williams, L., *Principal Component Analysis.* WIREs Computational Statistics, 2010. **2**: p. 433-459.

10. Wold, S., K. Esbensen, and P. Geladi, *Principal Component Analysis.* Chemometrics and Intelligent Laboratory Systems, 1987. **2**(1-3): p. 37-52.

11. Trygg, J., E. Holmes, and T. Lundstedt, *Chemometrics in Metabonomics.* Journal of Proteome Research, 2007. **6**(2): p. 469-79.

12. Keun, H.C., et al., *Analytical Reproducibility in (1)H NMR-Based Metabonomic Urinalysis.* Chemical Research in Toxicology, 2002. **15**(11): p. 1380-6.

13. Weckwerth, W. and K. Morgenthal, *Metabolomics: From Pattern Recognition to Biological Interpretation.* Drug Discovery Today, 2005. **10**(22): p. 1551-8.

14. Whelehan, O.P., et al., *Detection of Ovarian Cancer using Chemometric Analysis of Proteomic Profiles.* Chemometrics and Intelligent Laboratory Systems, 2006. **84**(1-2): p. 82-87.

15. Wiklund, S., et al., *Visualization of GC/TOF-MS-Based Metabolomics Data for Identification of Biochemically Interesting Compounds using OPLS Class Models.* Analytical Chemistry, 2008. **80**(1): p. 115-22.

16. Wold, S., et al., *Orthogonal Signal Correction of Near-Infrared Spectra.* Chemometrics and Intelligent Laboratory Systems, 1998. **44**(1-2): p. 175-185.

17. Nicholson, J.K., et al., *Metabonomics: A Platform for Studying Drug Toxicity and Gene Function.* Nature Reviews. Drug Discovery, 2002. **1**(2): p. 153-61.

18. Forgue, P., et al., *NMR Metabolic Profiling of Aspergillus nidulans to Monitor Drug and Protein Activity.* Journal of Proteome Research, 2006. **5**(8): p. 1916-23.

19. Kaddurah-Daouk, R., B.S. Kristal, and R.M. Weinshilboum, *Metabolomics: A Global Biochemical Approach to Drug Response and Disease.* Annual Review of Pharmacology and Toxicology, 2008. **48**: p. 653-83.

20. Bergmann, F., H. Ungar-Waron, and H. Kwietny-Govrin, *Action of 8-Azaguanine and 8-Azaxanthine on Pseudomonas aeruginosa.* The Biochemical Journal, 1964. **91**(2): p. 270-6.

21. Norris, E.R. and A. Roush, *The Competitive Inhibition of Uricase Activity by 8-Azaguanine and by 8-Azaxanthine.* Archives of Biochemistry, 1950. **28**(3): p. 465-6.

22. Oestreicher, N., H.M. Sealy-Lewis, and C. Scazzocchio, *Characterisation, Cloning and Integrative Properties of the Gene Encoding Urate Oxidase in Aspergillus nidulans.* Gene, 1993. **132**(2): p. 185-92.

23. Kang, J., et al., *Application of a 1H Nuclear Magnetic Resonance (NMR) Metabolomics Approach Combined with Orthogonal Projections to Latent Structure-Discriminant Analysis as an Efficient Tool for Discriminating Between Korean and Chinese Herbal Medicines.* Journal of Agricultural and Food Chemistry, 2008. **56**(24): p. 11589-95.

24. Goodacre, R., et al., *Proposed Minimum Reporting Standards for Data Analysis in Metabolomics.* Metabolomics : Official Journal of the Metabolomic Society, 2007. **3**(3): p. 231-241.

25. Sinkov, N.A. and J.J. Harynuk, *Cluster Resolution: A Metric for Automated, Objective and Optimized Feature Selection in Chemometric Modeling.* Talanta, 2011. **83**(4): p. 1079-1087.

26. S., W., *Multivariate Data Analysis for Omics.* 2008: p. 1-228.

27. Eriksson, L., Johansson, E., Kettaneh-Wold, N., Trygg, J., Wikstrom, C., Wold, S., ed. *Multi- and Megavariate Data Analysis Part I: Basic Principles and Applications*. 2 ed. 2006, Umetrics.

28. Bro, R., et al., *Cross-Validation of Component Models: A Critical Look at Current Methods.* Analytical and Bioanalytical Chemistry, 2008. **390**(5): p. 1241-51.

29. Arlot, S., Celisse, A., *A Survey of Cross-Validation Procedures for Model Selection.* Statistics Surveys, 2010. **4**: p. 40-79.

30. Henningsson, M., et al., *PLS Model Building: A Multivariate Approach to Personality Test Data.* Scandinavian Journal of Psychology, 2001. **42**(5): p. 399-409.

31. Eriksson, L., J. Trygg, and S. Wold, *CV-ANOVA for Significance Testing of PLS and OPLS (R) Models.* Journal of Chemometrics, 2008. **22**(11-12): p. 594-600.

32. Werth, M.T., et al., *Analysis of Metabolomic PCA Data Using Tree Diagrams.* Analytical Biochemistry, 2010. **399**(1): p. 58-63.

33. Worley, B., S. Halouska, and R. Powers, *Utilities for Quantifying Separation in PCA/PLS-DA Scores Plots.* Analytical Biochemistry, 2013. **433**(2): p. 102-4.

34. Fendt, S.M., et al., *Tradeoff Between Enzyme and Metabolite Efficiency Maintains Metabolic Homeostasis Upon Perturbations in Enzyme Capacity.* Molecular Systems Biology, 2010. **6**: p. 356.

CHAPTER 3

REVISITING PROTOCOLS FOR THE NMR ANALYSIS OF BACTERIAL METABOLOMES[§]

3.1 Introduction

Metabolomics is the study of small molecules in a biological system that participates in the metabolic reactions responsible for cell growth, survival, and other normal cellular functions [1-3]. Additionally, the metabolome responds to transcriptional and translational alterations associated with genotypical, epigenetic, or environmental perturbations [4-7]. Thus, metabolomics provides an assessment of global perturbations with respect to an altered genome, proteome, or environment [2, 8, 9]. The simultaneous integration of genomic, transcriptomic and proteomic data has enabled an in-depth analysis of the interplay, interaction, and regulation of DNA, RNA and proteins [10-12]. Along this line, monitoring the bacterial metabolome and integrating the results with other "omics" data has provided valuable insights into bacterial adaptability [13], biofilms [14], evolution [15], pathogenesis [16], and drug resistance [17].

Depending on the organism and growth state, the total number of metabolites within a cell varies between several hundred to a few thousand, with a corresponding diversity in physical and chemical properties, such as size, stability, and concentration [18]. In addition to the challenge of the simultaneous study of all the metabolites within a

[§] Chapter 3 was adapted from Halouska.S., *et. al*., Revisiting Protocols for the NMR Analysis of Bacterial Metabolomes *Journal of Integrated OMICS* (2013). Reprinted with permission, copyright Proteomass 2013.

given biological system [19], the selection of an analytical technique will influence which metabolites are observed. NMR and MS are commonly employed for metabolomics, where both instruments can be interfaced with LC, GC, and CE systems to select and emphasize specific components of the metabolome [20-24]. NMR has a number of advantages in analyzing the metabolome that includes minimal sample handling and that it is not reliant on chromatography to purify or separate metabolites. In addition, multiple resonances from a single molecule increase the accuracy of metabolite identification and quantitation. This accuracy can be further enhanced by the application of ^{13}C- and ^{15}N-isotope labeling to enhance specific regions of the metabolome [25, 26]. Importantly, the choice of ^{13}C- or ^{15}N-labeled metabolite determines the region of the metabolome observed by NMR, providing significant flexibility in experimental design. In contrast to MS, NMR is a relatively insensitive technique and only observes the most abundant ($\geq$ 1 to 5 μM) metabolites. In addition, MS has the advantage of detecting a wider-range of the metabolome. However, because of the relatively low molecular-weight range of the metabolome, MS methods generally require chromatography to separate metabolites before analysis [27]. Additionally, variations in ionization and the occurrence of ion suppression in a complex mixture add uncertainty in detecting specific metabolites by MS [28]. Finally, quantitation by MS is typically more challenging than NMR. Taken together, NMR and MS each have strengths and weaknesses but should be viewed as complementary techniques [29].

NMR-based metabolomics have been used to study a wide range of biological systems such as tissues [30], biofluids [31], mammalian cell cultures [32], plants [33] and

bacteria [34-36]. The overall procedure for an NMR-based metabolomics study includes the following general steps: cell growth and harvesting, metabolite extraction, NMR data collection and analysis, multivariate statistical analysis, metabolite identification and quantification [37]. Typically, one-dimensional (1D) ^{1}H NMR spectra are used for a multivariate analysis such as principal component analysis (PCA) or orthogonal partial least squares projections discriminant analysis (OPLS-DA) [38, 39]. Both PCA and OPLS-DA provide global profiles of metabolome changes [40, 41]. Two-dimensional (2D) ^{1}H,^{13}C Heteronuclear Single Quantum Coherence (HSQC) or ^{1}H,^{1}H TOtal Correlated SpectroscopY (TOCSY) NMR experiments are used for the quantitative assessment of metabolite changes resulting from genetic modification or external stimuli [5, 14]. The ability to generate global profiles and quantitative differences coupled with the ease of applying NMR-based metabolomics has contributed to the rapid growth of the NMR metabolomics field. While NMR data acquisition and analysis methods are improving, care must be taken to ensure that the methods are appropriate to the task at hand and generate biologically relevant information. As an example, protocols to efficiently extract metabolites without inducing cellular changes are essential for success [32, 42]. In brief, the observed changes in the metabolome should reflect a change in the state of the system instead of how the cells are handled and processed. Similarly, variations in instrument performance, choice of procedures for data collection and processing, and invalidated models from multivariate analysis may induce unintended biases or incorrect interpretation of metabolomics data [43-46].

Since NMR-based metabolomics is a relatively new and still developing technology, improving and enhancing the experimental protocols is necessary to advance the field and ensure continued success. Toward this end, we describe our recently developed and optimized protocols for the application of NMR metabolomics to microbial samples. We present our current methodology and also discuss the challenges associated with each major step of the process: (i) sample preparation, (ii) NMR data collection and processing, (iii) multivariate statistical analysis, (vi) metabolite identification and network generation. Specifically, the overall methodology will be discussed in detail, where a number of key features will also be highlighted, such as automation, bioinformatics, experimental design, and harvesting the metabolome. The focus of our efforts has been to identify and minimize procedural steps that negatively influence the outcome of an NMR-based metabolomics experiment.

The development of the metabolomics protocol was a group effort where Steven Halouska and Bo Zhang demonstrated significant and equal contributions to the project. Steven Halouska was responsible for the development and applications for the overall experimental design of the NMR metabolomics methods. This includes developing methods for both metabolomics fingerprint analysis and metabolomics profiling. Bo Zhang demonstrated significant contributions to the development and optimization of the bacterial sample handling used for both fingerprint analysis and metabolomics profiling. This includes optimizing the strategy for sample growth, cell quenching, and extraction. Both Steven Halouska and Bo Zhang had equal contribution in the optimization in the data analysis.

3.2 Experimental Design

A general protocol for the analysis of bacterial metabolomes using NMR is shown in Figure 3.1. The flow diagram illustrates procedures for both a global analysis of metabolome changes (metabolomics fingerprinting); and the identification and quantitation of specific metabolites correlated with the biological process (metabolomics profiling). The overall process consists of the following key steps: bacterial cultivation and harvesting, metabolite extraction, NMR data collection and analysis, multivariate statistical analysis, metabolite identification and quantification. Successful metabolomics sample preparation involves three steps: The first step is the simultaneous growth of all of the bacterial cultures or as many as is practical at a time. The bacteria are grown in a standard medium for fingerprint analysis, whereas the medium is supplemented with a ^{13}C-labeled metabolite for metabolomics profiling [47, 48]. After the bacteria are grown for a defined time or they have achieved a specified cell density, bacteria are harvested and quenched to halt all enzymatic processes and washed to remove the medium. The third sample preparation step involves lysing the cells and extracting the metabolome. A variety of solvents are routinely employed depending on the solubility of the targeted metabolites (cytosolic metabolites, lipids, *etc*.). The metabolomics samples are then used to generate a series of NMR spectra, which are used for the multivariate statistical analysis, metabolite identification and quantification. The individual steps of the NMR-based metabolomics protocol will be discussed in detail highlighting challenges associated with each step.

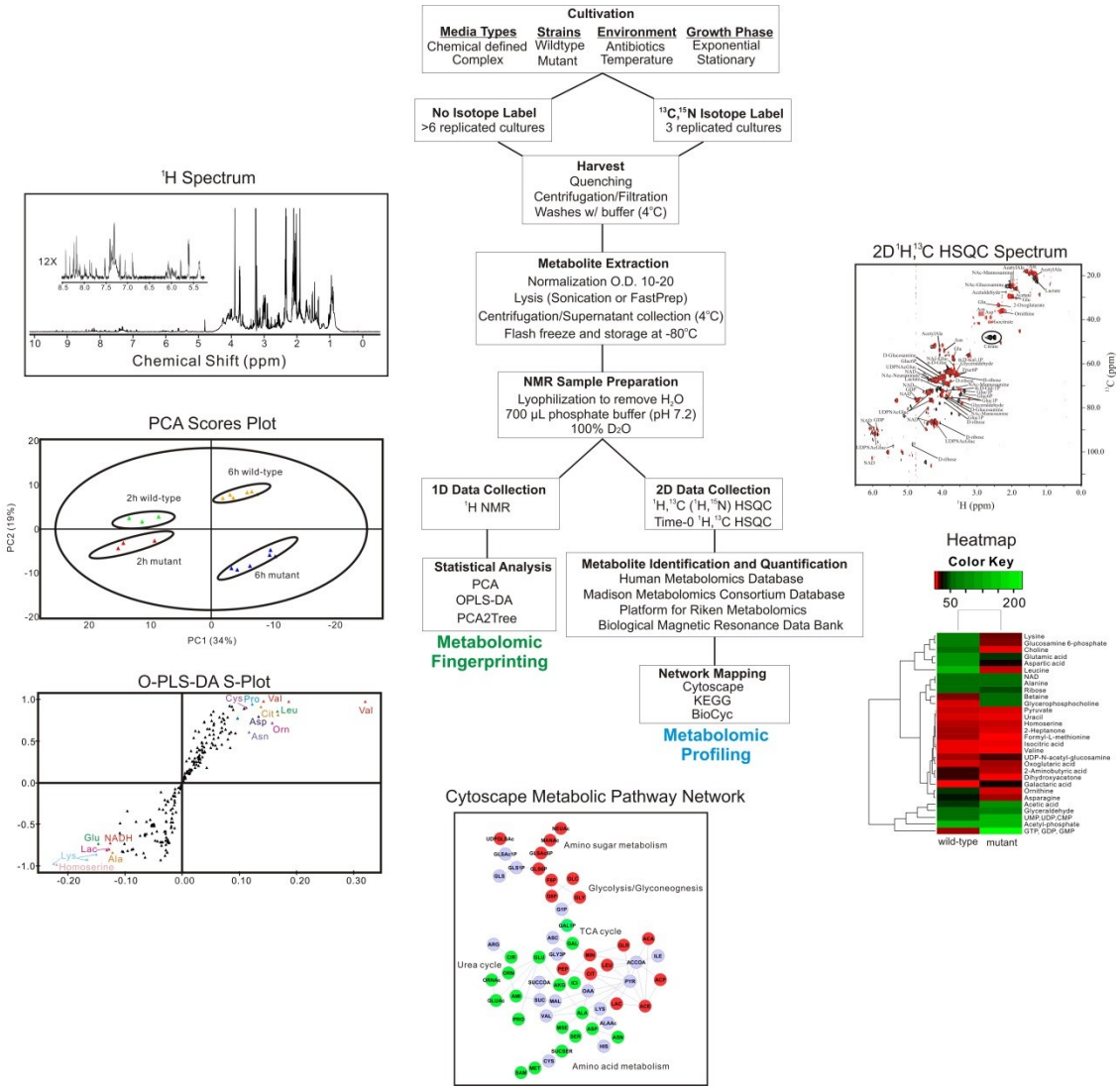


Figure 3.1. A flow chart of our protocol used for the NMR analysis of bacterial metabolomes.

3.2.1 Identification of an appropriate biological system for a metabolomics study

NMR-based metabolomics is an important tool in systems biology research. The quantitative and qualitative measurement of metabolites from cytosolic extracts can be extremely valuable for investigating cellular processes, pathogenesis, and the effects of drugs or the environment on bacteria. Unfortunately, the bacterial metabolome is a complex mixture of metabolites and numerous interconnected metabolic and signaling pathways. This high interconnectivity may result in significant metabolite concentrations changes far from the origin of the perturbation (inhibited, inactivated or down-regulated protein). Correspondingly, it is easier to observe changes to the metabolome than deduce the primary source of the perturbation after its impact has rippled throughout the metabolome. As an illustration, treating a bacterial culture with a particular drug would be expected to lead to a global change in the metabolome, but interpreting these changes to identify the therapeutic target is extremely challenging. To address this challenge, the *in vivo* mechanism of action of a potential drug lead may be determined by comparing these metabolome changes to other drugs with known biological targets [49] or to a mutant bacteria where a specific protein target is ablated or modified by genetic inactivation [50, 51]. This example illustrates that the comparative analysis between two or more metabolomes is an effective application of metabolomics. In order to obtain reliable insights into the physiology of bacteria or any other organism, it is essential to identify and establish at least two reference metabolomes (wild-type vs. mutant, drug-resistant vs. drug susceptible, nutrient-rich vs. nutrient-limited, *etc*.) for a comparative analysis. Once the reference conditions are established, bacteria can be exposed to any

range of experimental variables such as a drug treatment, environmental stimuli (pH, temperature, nutrient change), or gene knockout (mutants, RNAi, inhibitor) to determine if similarities exists with the reference metabolome. The similarity between metabolomes infers an overlap in the underlying physiological processes or responses that gave rise to the metabolome changes. We have used this approach to demonstrate the similarity of *Staphylococcus epidermidis* metabolomes resulting from exposure to divergent environmental stressors that are known to facilitate biofilm formation [5, 14]. Our results suggested that the tricarboxylic acid (TCA) cycle acts as a metabolic signaling network to transduce external stresses into internal metabolic signals. This conclusion was only possible because the experimental design was based on comparing the metabolomes of the *S. epidermidis* wild-type strain 1457 and an aconitase mutant strain 1457-*acnA::tetM* with and without the treatment of biofilm stressors. In summary, the successful outcome of a metabolomics study hinges on the experimental design and the proper choice of the cellular metabolomes to be compared.

3.2.2 Minimization of unintended bias and biologically irrelevant variations

In addition to the proper choice of bacterial strains to compare in a metabolomics study, the experimental protocols must be optimized to reduce unwanted variation or bias in the collection of cell-free lysates. It is essential to ensure that any metabolome changes are limited to biologically relevant factors and are not caused by the handling or processing of the samples. Thus, the key to metabolomics is establishing an efficient methodology that closely captures the true state of the metabolome [52]. Fundamental to

a successful metabolomics experiment is maximizing the uniformity of the preparation, handling, processing, and analysis of each replicate sample [35, 45, 53-56]. In instances where cultivation and/or processing variation is unavoidable (*e.g.*, if multiple incubators are required to accommodate all the replicates), then the cultures should be randomly distributed between the incubators to minimize bias. Ideally, all of the metabolomics samples should be handled by the same person because subtle differences in individual techniques may influence the outcome. If multiple investigators are required to efficiently handle the samples, each researcher should be assigned a specific set of tasks that are consistently applied to each sample. For example, one investigator lyses all the bacterial cells while another performs the metabolome extraction procedure on every sample.

As with sample cultivation and preparation methods, the NMR spectra generated from metabolomics samples need to accurately represent the state of the system. In other words, the NMR spectra must reflect the actual concentrations and identity of the metabolites present in the biological sample at the time of harvest. If the sample preparation and data acquisition represent the metabolic status at the time of harvest, then multivariate statistical techniques, such as PCA and OPLS-DA, will enhance the identification of similarities or differences in the NMR spectra, and, correspondingly, between the bacterial metabolomes [39]. These multivariate statistical techniques typically involve multiple replicates of 1D ^{1}H NMR or 2D ^{1}H,^{13}C HSQC spectra for each bacterial class or group (*e.g.* wild-type, mutant, drug treated, *etc*.). The exact number of replicates is dependent on a number of factors: (i) the variance within a group, (ii) the

variance between groups, (iii) the number of variables, and (iv) type of statistical analysis performed [57, 58].

In most metabolomics experiments, the number of biological samples is significantly smaller than the number of variables; in this case, the variables correspond to peaks in the NMR spectra or the detectable metabolites [58]. For this reason, a larger number of replicates (≥ 6) per class are required to obtain a statistically significant PCA or OPLS-DA model. While greater numbers of replicates are desirable, there are practical considerations to increasing the number of replicates, including increased experimental time, availability of incubator space, and practical limits on the number of samples that can be simultaneously prepared and processed within a reasonable time frame. The increased time, larger number of samples, and added complexity may be detrimental to maintaining consistency between samples, where metabolite stability may become more of an issue [59]. So the potential benefit in improving the reliability of the PCA or OPLS-DA models may be negated by too large of a sample size if sample consistency is sacrificed. In general, 6 to 10 replicates per class can be routinely handled while providing a statistically significant PCA or OPLS-DA model. Lastly, to increase the sample consistency, the application of an automated sample changer or flow-probe can minimize variability by eliminating human involvement and providing a uniform and consistent protocol for NMR data collection. Nevertheless, instrument drift may still occur during the high-throughput experiment so it is also important to randomize the samples during data collection.

3.3 Sample Preparation

3.3.1 Bacterial cultivation

Consistency is critical to metabolomics, where variations in a bacterial metabolome may be introduced by cultivation protocols. To achieve the reproducible cultivation of bacteria requires consideration of three variables: bacterial strain, culture medium, and cultivation conditions. Strain selection is often driven by investigator preference, availability, or cultivability. The choice of culture medium will largely depend on which, if any, isotopically-labeled metabolite is being followed. For example, when using ^{15}N-arginine, it is impractical to add labeled arginine to a complex medium containing an unknown concentration of unlabeled arginine. In this example, to achieve maximal labeling of the bacteria, it would be best to use a chemically defined medium lacking arginine. Importantly, the culture medium has to be consistently employed throughout the metabolomics study. A different culture medium cannot be used for metabolomics fingerprinting and profiling, it cannot vary based on the requirements of the bacterial strain or to accommodate an experimental variable. Different culture media will induce changes in the metabolome that will mask or complicate any analysis. Bacterial cultures also need to be properly handled in order to avoid inducing biologically irrelevant changes. For example, pre-warming the culture medium prevents temperature shock and minimizes variation between biological replicates. Similarly, randomizing the samples from each group and class also minimizes bias that may occur if all the samples are processed in a predefined order. Importantly, different cell types may require special care or different handling protocols. Cultivation conditions will also vary depending upon

the experiment; however, consideration must be given to each of the following: temperature, pH buffering (if used), % CO_2 (if used), the flask-to-medium ratio, the revolutions per minute of agitation (if used), the use of baffled or non-baffled flasks, and the inoculum dose. In effect, one protocol does not necessarily "fit all" and a general metabolomics protocol needs to be optimized for each experiment and cell type.

Deciding on the number of bacterial cultures needed for a metabolomics study and identifying the optimal culture size are important next steps. The volume of the bacterial culture should be large enough to provide a sufficient number of cells to maximize the NMR signal-to-noise, but small enough to simplify the handling of numerous replicate samples. An appropriate cell density must be determined empirically for each species and bacterial strain, which will also limit the culture size. Similarly, the growth phase chosen for harvesting bacteria will also contribute to defining the optimal culture size since cell density changes drastically between the lag, exponential and post-exponential phases. In our experience with staphylococcal and mycobacteria cultures, media volumes between 15 to 50 mL are used to grow cells to an optical density at 600 nm ($O.D._{600}$) of 1-2 for bacterial cultures collected during the exponential phase. Conversely, media volumes of between 3 to 5 mL are used to grow cells to an $O.D._{600}$ of 3 to 7 for bacterial cultures collected during the post-exponential phase (e.g., 6 to 7 for *Staphylococcus epidermidis*, and 3 to 4 for *Mycobacterium smegmatis*). The overall goal is to have an $O.D._{600}$ of 10 to 20 after the bacterial cells have been concentrated to a final volume of 1 ml. This will ensure metabolite concentrations sufficient for detection by NMR. These culture volumes and $O.D._{600}$ values should be viewed as guidelines and

targeted goals that may require further optimization for different bacterial strains or species. As previously stated, the number of bacterial cultures determines the statistical significance of class differentiation or metabolite changes. In our experience, ten replicates are an optimal choice for metabolomics fingerprinting and only three replicates are needed for metabolomics profiling.

Metabolomics profiling requires ^{13}C- or ^{15}N-labeled metabolites and defines the choice of culture media. In our laboratories, we typically label staphylococci using ^{13}C-glucose in the complex medium tryptic soy broth (TSB) that is devoid of unlabeled glucose [6, 7]. This medium allows for maximal biomass generation, while assuring that nearly all (~99%; 1.1% is due to naturally occurring ^{13}C) of the ^{13}C-labeled metabolites in the metabolome were derived from glucose. Similarly, we have labeled mycobacteria using ^{13}C-glucose or ^{13}C-glycerol in Middlebrook 7H9 Albumin Dextrose Complex (MADC; Becton-Dickinson) media. We have also supplemented culture media with ^{13}C-alanine, ^{13}C-aspartate, ^{13}C-glutamate, ^{13}C-proline and ^{13}C-pyruvate as a more targeted approach to the analysis of the metabolome. These metabolites are associated with a limited number of metabolic pathways. The analysis of the metabolome can be further focused by using a targeted metabolite where only one or a few specific carbons in the metabolite are ^{13}C labeled. Only the metabolic pathways involving the specific ^{13}C-labeled carbon will be observable by NMR. The concentration of the ^{15}N-, or ^{13}C-labeled metabolite needs to be high enough ($\geq$ 1 to 5 μM) to be detected by NMR. In our experience with staphylococcal and mycobacterial cultures, the volumes range from 25 mL to 100 mL and the culture media should be supplemented with approximately 2.5 to 4

g/L of $^{13}C_6$-glucose or ~10-15 mg/L of a targeted metabolite like ^{13}C-D-alanine in order to acquire a 2D ^{1}H-^{13}C HSQC spectrum with acceptable signal-to-noise.

Ideally, each bacterial culture should contain the same number of cells and be at the same growth phase when harvested. In reality, differences in cultivation conditions, media, and/or bacterial strains may substantially affect growth rates and/or growth yields. The two more common approaches to compensate for different bacterial growth rates are: collect the bacteria when they have reached the same cell density, but at different times to account for the different growth rates; and harvest the bacteria at the same time but harvest equivalent cell numbers. As examples, in *staphylococci*, the exponential and post-exponential growth phases were typically analyzed at the 2 h and 6 h time points, respectively [5]. For our mycobacterial experiments, a consistent growth phase was achieved by harvesting bacteria at a uniform $O.D._{600}$ of 1.2. In practice, it is extremely difficult to harvest every bacterial culture with an identical $O.D._{600}$ value. To correct for this variability, all the bacterial cultures are normalized to the same number of cells. Simply, the cultures are suspended into a phosphate buffer until the $O.D._{600}$ values are equal. Typically the cultures are concentrated to an $O.D._{600}$ of 10 to 20 into a final volume of 1 mL prior to lysis. Alternatively, the bacterial cell cultures can be normalized based on colony-forming units (CFU) or total protein concentration.

To ensure consistency, the experimental variable such as a drug treatment, environmental stimuli, or gene knockout needs to be uniformly applied to the "treatment" class. An additional consideration for treatment of cultures, is that the impact on the metabolome should be strong enough to detect [49]. In other words, a particular drug

dosage needs to be large enough to affect the cellular metabolome relative to untreated cells, but should not induce cell death. In our experience, a drug concentration that inhibits bacterial growth by 50% relative to untreated cultures is a desirable target [49, 50]. The availability from the literature of a minimal inhibitory concentration for the strain (MIC), or otherwise for the population isolates (MIC_{50}), provides a good starting point for optimizing a drug dosage, but the actual dosage must be determined empirically for each set of cultivation conditions. In our experience, literature MIC or MIC_{50} values tend to be too low for cultivation conditions used for metabolomics. We typically test drug concentration ranges at between 1 to 24 times the reported MIC or MIC_{50} values in order to identify an optimal drug dosage. Importantly, this also implies that drugs with a range of biological activity will require different drug concentrations in a metabolomics study; hence, the use of the 50% inhibition of growth is used as a metric as opposed to drug concentration. Typically, in our experiments the drug treatments are normally administered during the exponential phase and the bacteria are allowed to grow for at least one generation before harvesting. In our experience, this provides a sufficient amount of time for the drug to affect cell physiology and induce a perturbation in the metabolome. Administering a drug at an earlier time point can be problematic because of the inability to harvest enough bacteria.

3.3.2 Quenching, washing and harvesting the bacterial cells

Speed is critical to harvesting bacteria without inducing a change to the metabolome. Changes occur quickly because of different metabolite turnover-rates,

varying stabilities, and the induction of stress responses, among other factors [59-61]. As bacteria are being harvested, the environment is changing dramatically: (i) the bacteria are either adhered to the surface of filter paper or at the bottom of a centrifuge tube under anaerobic conditions, (ii) the temperature is changed from 37^{o}C to ~ 0^{o} C, and (iii) the growth media is replaced with either double distilled water or a phosphate buffer. To prevent perturbations to the metabolome caused by handling of the cell samples, the bacteria need to be rapidly quenched in order to stop all cellular processes from responding to these changes. Quenching efficiency has been widely discussed in the literature [42, 62-64]. Importantly, the quenching technique employed also defines the washing protocol and the order that quenching, washing and cell separation takes place. Our quenching techniques consist of a filtered cells being quickly submerged into liquid nitrogen or the cells and media being directly added to -60^{o}C cold ethanol or methanol solution while being vortexed. The media and ethanol/methanol volumes are at an equal 1:1 ratio. After centrifugation, the supernatant is decanted and disposed of, and the cell pellet is ready for washing. Unfortunately, there is a possibility of cell leakage and loss of metabolites when the cells are directly added to the cold ethanol or methanol solution.

Before intracellular metabolites can be analyzed, the bacteria need to be rapidly separated from the culture media. Filtration and centrifugation are both routinely used in our laboratory to separate bacterial cells from the media. Filtration has a definitive advantage because it is significantly faster than centrifugation, but challenges in removing and collecting intact cells from filter paper may lead to sample variability. Conversely, the variability between metabolome replicates is expected to be reduced with

centrifugation because of the ease in handling the cells. Nevertheless, our experience with washing bacterial cells using either filtration or centrifugation has resulted in essentially identical metabolomics fingerprints (Figure 3.2a); thus, any undesirable variation within a group likely occurs during sample preparation. Metabolome differences between replicate bacterial cultures likely arises from a combination of subtle variations in the number of cells, culture conditions, length of bacterial cultivation time, and sample handling.

The use of centrifugation or filtration also determines the quenching protocol [36]. Harvesting bacteria using centrifugation requires quenching the bacteria using the direct addition to -60^{o}C cold ethanol or methanol. The bacteria, culture media, and quenching solution are in a properly sized conical centrifuge tube that is centrifuged for 8 minutes at 4,284 *g* (bucket rotor). Following centrifugation, the culture media and quenching solution are decanted and the bacteria are suspended in 30 mL of an ice cold wash. We routinely wash bacteria with ice cold double distilled water, phosphate buffer (20 mM, pH 7.2), or phosphate buffered saline (PBS; 6 mM phosphate buffer, pH 7.4, 137 mM NaCl and 2.7 mM KCl) to remove residual media and avoid contamination of the metabolome. The bacteria are centrifuged again, the wash is decanted off and the process is repeated. After two washes, the cell pellet is suspended in 1 mL of the ice cold wash and transferred to a 2 mL vial for cell lysing. Additional washings provide an insignificant benefit in removing media contaminates, but results in an undesirable increase in time. Also, the buffered wash eliminates any impact on the cells from a pH change, but double distilled water eliminates the possibility of "salting-out" any

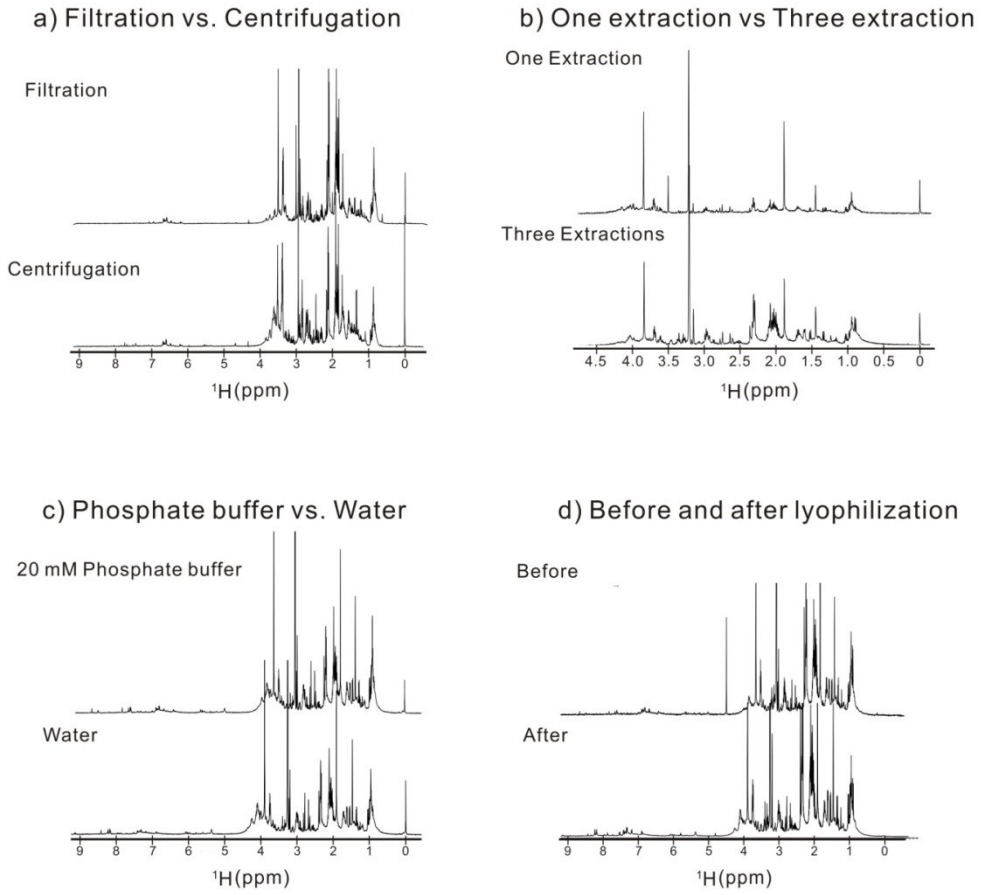


Figure 3.2. Illustrations of the impact of **a)** filtration and centrifugation, **b)** number of extraction steps, **c)** type of wash buffer, and **d)** lyophilization on the composition of the metabolome.

metabolites that may occur from concentrating the buffer in subsequent steps [65]. The cells are kept on ice throughout this entire process.

Harvesting bacteria by vacuum filtration collects the bacteria on sterile filter paper (0.45 μm pore size; Millipore), while simultaneously removing the media. The number of bacteria that can be harvested onto a filter must be empirically determined to prevent a filter blockage. Under proper conditions, removing the media should take less than a minute, and should never exceed two minutes. If this cannot be achieved, then the bacteria need to be harvested using centrifugation. After filtration, the filter paper containing the cells is then quickly placed into a 50 mL conical centrifuge tube and submerged into liquid nitrogen to freeze and quench the cells. The conical vial is then warmed by placing it into a bucket of ice for ~1 to 2 minutes. This prevents freezing of the 1 mL of wash that is added to the conical vial. The cells are gently removed from the filter paper with the wash and then transferred to a 1.5 mL microcentrifuge tube. The cells are centrifuged and washed twice (1 mL) as before.

3.3.3 Cell lysing and metabolite extraction

The cells need to be lysed in order to extract the cellular metabolome. Cells can be lysed by chemical or physical means, but the use of chemicals runs the added risk of contaminating the metabolome. Thus, the FAST-Prep bead beating method of lysing cells is our preferred approach. Before lysing, the samples are normalized to an $O.D._{600}$ value of between 10 to 20 units. This insures an equivalent number of cells, comparable metabolite concentrations, and a detectable NMR spectrum. Each sample is placed into a

2 mL micro-centrifuge tube with small glass beads (Lysinf Matrix B; MP Biomedical) and 1 mL of extraction buffer. The cells are crushed by bead beating for 40 to 60 seconds in the FAST-Prep instrument at a speed of 6.0 m/s. The sample is then centrifuged for 2 minutes at 17,000 g to pellet the cell debris. The supernatant with the extracted metabolites is collected. The cell debris is washed 1 to 3 times with 1 mL of the extraction buffer to maximize the metabolome yield (Figure 3.2b). Also, double distilled water or a phosphate buffer are routinely used as the extraction buffer, since both approaches provide similar results (Figure 3.2c). All extracts per sample are combined for lyophilization, where the sample is then dissolved in 700 μL of a phosphate buffer in D_2O at pH 7.2 (uncorrected). Lyophilization may negatively impact some volatile metabolites, but, in general, no effect is observed (Figure 3.2d). A major concern during the extraction step is maximizing the overall yield while minimizing any perturbation to the metabolome. In our experience, the cell lysing and metabolite extraction process will require approximately 45 minutes for 30 cultures. The metabolomics samples can be stored in a -80^{o}C freezer or directly lyophilized overnight.

3.4 NMR Spectroscopy

3.4.1 One-dimensional ^{1}H NMR methodology

One-dimensional (1D) ^{1}H (proton) NMR is an unbiased, nonselective, and nondestructive approach that requires no modification of the samples, where the data can be collected in a high-throughput manner. A 1D ^{1}H NMR spectrum contains numerous proton signals generated from a complex metabolomics mixture, where the chemical shift

of each signal describes the structural characteristic of a specific metabolite. Moreover, the peak intensities or volumes are directly proportional to the concentration of each metabolite. Quantification of metabolites can be achieved by using an internal standard with a known concentration, where we routinely use 50 μM 3-(trimethylsilyl) propionic acid-2,2,3,3-d4 (TMSP-d4, Sigma). Thus, 1D ^{1}H NMR experiments combined with multivariate statistics are commonly used for the global analysis of the metabolome.

Collecting 1D ^{1}H NMR data for metabolomics is fast and simple, and provides highly reproducible and accurate results. Importantly, the NMR experimental parameters need to be identical for each metabolomics sample in order to collect reliable metabolomics data. Any per sample variation will erroneously bias the resulting clustering patterns in the PCA and OPLS-DA scores plot. To avoid this and maintain sample consistency, we use a BACS-120 sample changer, Bruker ICON-NMR, an automatic tuning and matching (ATM) unit, and autoshim to automate the NMR data collection. Nevertheless, instrument drift may still occur during the high-throughput metabolomics screen, so it is also important to randomize the samples during NMR data collection. If an NMR spectrum is collected first for all the control samples followed subsequently by each treatment class, there is a significant potential of inducing a biologically irrelevant bias into the analysis. The clustering pattern in the PCA and OPLS-DA scores plot may be dominated by the order of data collection instead of the expected biological differences.

In our laboratory, a typical 1D ^{1}H NMR spectrum is collected using 128 scans and 32k data points on a Bruker 500 MHz Avance DRX NMR spectrometer equipped with a

triple-resonance, Z-axis gradient cryoprobe. The acquisition time is approximately 10 minutes per sample. The goal is to obtain optimal signal to noise while minimizing the total experimental time. We previously demonstrated that spectral noise is detrimental to the resulting PCA and OPLS-DA scores plot [53]. Random noise fluctuations results in large and irrelevant variations in the scores clustering. To avoid this problem, spectral noise needs to be removed prior to PCA and OPLS-DA. Correspondingly, the quality of the within class clustering in PCA and OPLS-DA scores plot is directly dependent on the spectral signal-to-noise (Figure 3.3). The within class variance decreases dramatically as the number of scans (signal-to-noise) is increased from right to left in the scores plot. Importantly, the spectral noise was still removed prior to PCA. Thus, the accuracy of identifying similarities or differences between multiple classes is dramatically improved by reducing within class variance, which is achieved by improving spectral sensitivity. Also, correctly identifying class differences improves with the number of replicates (Figure 3.4). The statistical significance of cluster separation as measured by *p*-values [66] is shown to decrease as both a function of group variance and the number of replicates. As a result, we prefer to use ten replicates per class and strive to achieve an average signal-to-noise ratio of > 100 to 200. This is achieved by simply increasing the number of scans or the number of cells, whichever is more practical. While signal-to-noise has a dramatic impact on scores clustering, PCA and OPLS-DA is indifferent to changes in spectral resolution unless the number of data points is ≤ 2K.

A D_2O phosphate buffer is the typical solvent of choice for aqueous metabolomics samples in order to efficiently remove residual water signals and avoid interference from

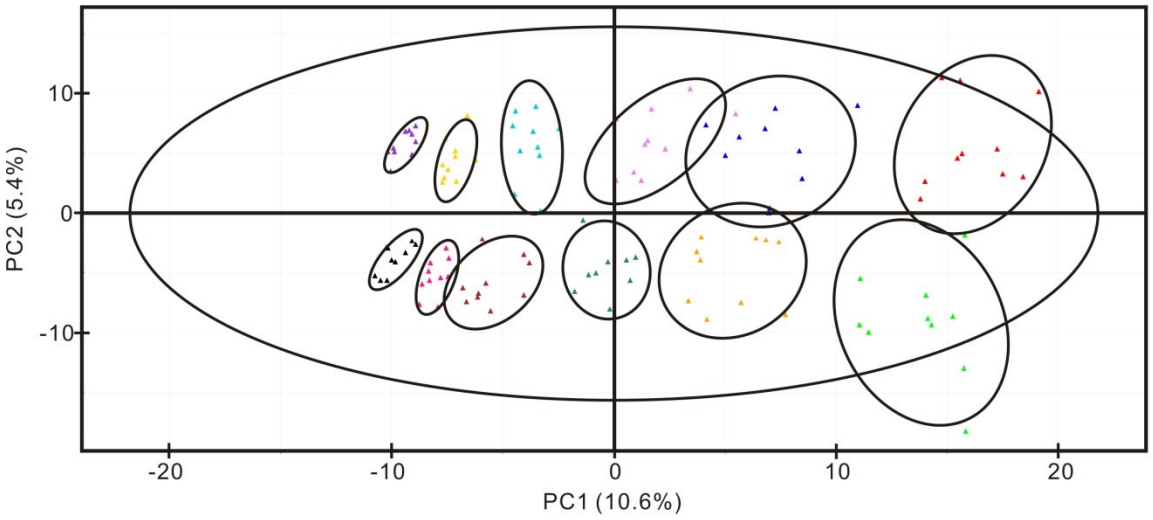


Figure 3.3. Illustration of the impact of the NMR signal-to-noise on within class variation in a PCA scores plot. From right to left, the 1D ^{1}H NMR spectra were collected with an increasing number of scans (1, 2, 4, 8, 16, and 32) resulting in a proportional increase in signal-to-noise. All other experimental parameters were kept constant.

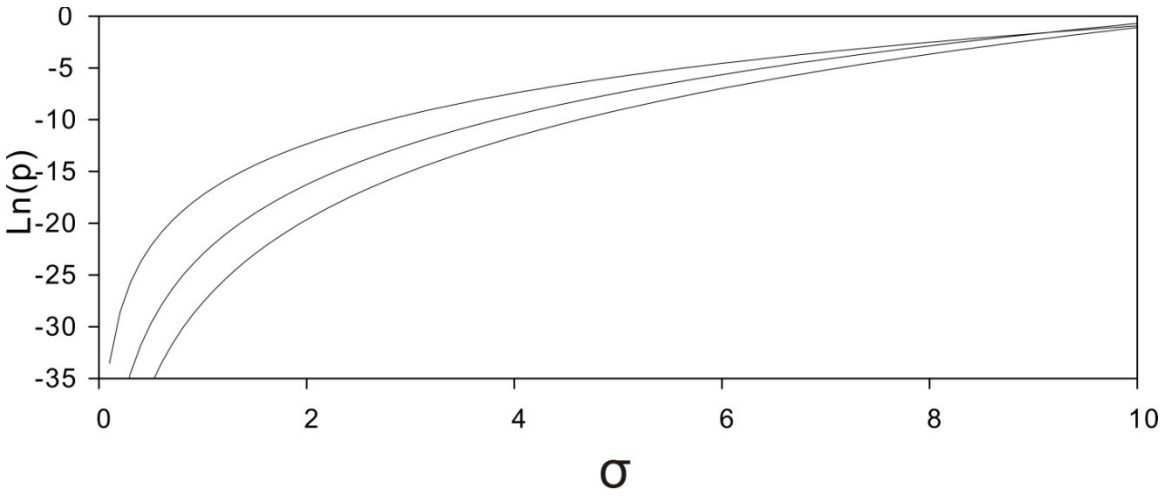


Figure 3.4. Illustration of the impact of within group variation and the number of replicates on the *p* values calculated between clusters in a simulated PCA scores plot. From top to bottom, *p* values from the simulated PCA scores plot were calculated with an increasing number of replicates (6, 8, 10) resulting in a proportional decrease in *p* values. Similarly, increasing the group variation by increasing the standard deviation (σ) per cluster resulted in a significant increase in *p* values.

buffer signals. Water and buffer signals are problematic since they can distort the NMR spectrum and may overlap and obscure important metabolite signals. Most NMR processing software can automatically remove the residual water peak, but extra data processing is required to correct for baseline distortions induced by the solvent. Unfortunately, simply applying a baseline correction changes the PCA and OPLS-DA clustering patterns [67]. Furthermore, different baseline correction protocols will induce variable changes into the scores plot. Also, removing the residual water peak may result in a potential loss of information by also removing metabolite peaks near the water signal. Instead, a water suppression technique that experimentally removes the water peak without inducing baseline distortions is the preferred alternative.

There are a variety of NMR pulse sequences for water suppression that are available to the metabolomics community, such as WATERGATE, water pre-saturation, WET, and PURGE [68-72]. Our preferred choice for a water suppression pulse sequence is Solvent-Optimized Gradient-Gradient Spectroscopy (SOGGY). SOGGY does an outstanding job in eliminating the water signal without inducing any base line distortions (Figure 3.2) [72]. SOGGY is a variant of excitation sculpting that employs a pulsed field gradient with a simple phase-alternating composite pulse. SOGGY offers the flexibility to optimize the 180 degree hard pulse to achieve optimal excitation of the water signal, and adjusting the 180 degree soft pulse to optimize the range of the water frequency to be suppressed [72, 73]. As a result, SOGGY efficiently suppresses the water signal while removing any phase cycle artifacts. A flat baseline is obtained while also maintaining

metabolite signals near the water signal [72]. SOGGY completely eliminates the need to apply any baseline correction.

3.4.2 Two-dimensional ^{1}H-^{13}C HSQC NMR methodology

The severe overlap of signals in a 1D ^{1}H NMR spectrum is a challenge for metabolite identification. The difficulty arises because hundreds to thousands of peaks occupy a small chemical shift range (~10 ppm), where multiple metabolites share similar chemical shifts. Thus, we typically do not use 1D ^{1}H NMR spectra to assign metabolites. Instead, we routinely use 2D ^{1}H-^{13}C HSQC experiments for metabolite assignments. The 2D ^{1}H-^{13}C HSQC experiment is a more reliable approach for metabolite identification because of the significantly higher resolution and the correlation between ^{1}H and ^{13}C chemical shifts for each C-H pair in a molecule [74, 75]. Also, the 2D ^{1}H-^{13}C HSQC experiment simplifies the analysis of the metabolome because only compounds containing a ^{13}C-carbon derived from the ^{13}C-labeled metabolite added to the media will be detected.

In our laboratory, we use a standard 2D ^{1}H-^{13}C HSQC pulse sequence on Bruker 500 MHz Avance DRX NMR spectrometer equipped with a triple-resonance, Z-axis gradient cryoprobe. An acceptable signal-to-noise is achievable using 64 scans. Similarly, a reasonable digital resolution is achieved by collecting 2K and 128 data points in the direct and indirect direction, respectively, with a corresponding spectral width of 10 ppm and 140 ppm along the ^{1}H and ^{13}C axis, respectively. Since some aromatic C-H pairs have a ^{13}C chemical shift greater than 140 ppm, the spectrum will contain folded peaks,

but the folded peaks will not interfere with or overlap with other metabolite peaks due to their unique position along the ^{1}H axis (~ 7.0 ppm). In general, the 2D ^{1}H-^{13}C HSQC experiment requires approximately 4 hours per sample on our system.

A conventional 2D ^{1}H-^{13}C HSQC spectrum is useful for detecting metabolite changes by overlaying multiple spectra to identify missing peaks or peaks with significant intensity changes. Unlike 1D ^{1}H NMR spectra, obtaining metabolite concentrations is more difficult because peak intensities are dependent on J-couplings, dynamics and relaxation, in addition to metabolite concentrations [76, 77]. To quantify absolute metabolite concentrations, we use the Time-Zero HSQC ($HSQC_0$) experiment [76]. This approach requires collecting a series of three HSQCs spectra ($HSQC_1$, $HSQC_2$, $HSQC_3$) with an increased number of pulse sequence repetitions. A natural log plot of peak areas or intensities versus the increment number (1,2,3) allows for an extrapolation back to increment 0 or zero-time. The experimental parameters used in the $HSQC_0$ experiment is similar to the conventional method, but with some minor variations. The number of scans is increased to 128 due to the decrease in signal-to-noise in $HSQC_2$ and $HSQC_3$. To partially compensate for the increase in experimental time, the number of data points in the indirect dimension is reduced to 64. In general, the $HSQC_0$ set of experiment requires approximately 6 hours per sample on our system.

3.5 Data analysis

3.5.1 Preprocessing of 1D ^{1}H NMR data

The 1D ^{1}H NMR spectra are minimally processed (Fourier transformed and phase corrected) using ACD/1D NMR Manager (Advanced Chemistry Development). Each NMR metabolomics sample contains 50 μM of TMSP-d_4 as an internal standard, where each NMR spectrum is referenced to the TMSP-d_4 peak and uniformly aligned to 0.00 ppm. Also, all peak heights are normalized to the intensity of the TMSP-d_4 peak. Intelligent bucketing within the ACD/1D NMR Manager is then used to integrate each spectral region with a bin size of 0.025 ppm. The spectra are normalized; noise regions and residual solvent and buffer resonances are removed, and then the remaining bins are scaled prior to PCA and OPLS-DA using SIMCA12.0+ (UMETRICS).

The need for data normalization and scaling prior to multivariate statistical analysis has been extensively discussed in the literature [78, 79]. Normalization adjusts for experimental variations between replicates, different number of cells, varying signal-to-noise, etc., and minimizes these contributions to the clustering patterns in PCA and OPLS-DA scores plot. We have encountered significant success in using a Z-score or center averaging the spectrum,

$$Z = \frac{X_i - \bar{X}}{\sigma} \tag{3.1}$$

where $\bar{X}$ is the average signal intensity in a given spectrum, σ is the standard deviation in the signal intensity, and X_i is the signal intensity within bin i (Figure 3.5a). After normalization, all the noise bins are uniformly removed. This was initially accomplished by manually identifying a “reference” noise region above 10 ppm and below 0 ppm; and

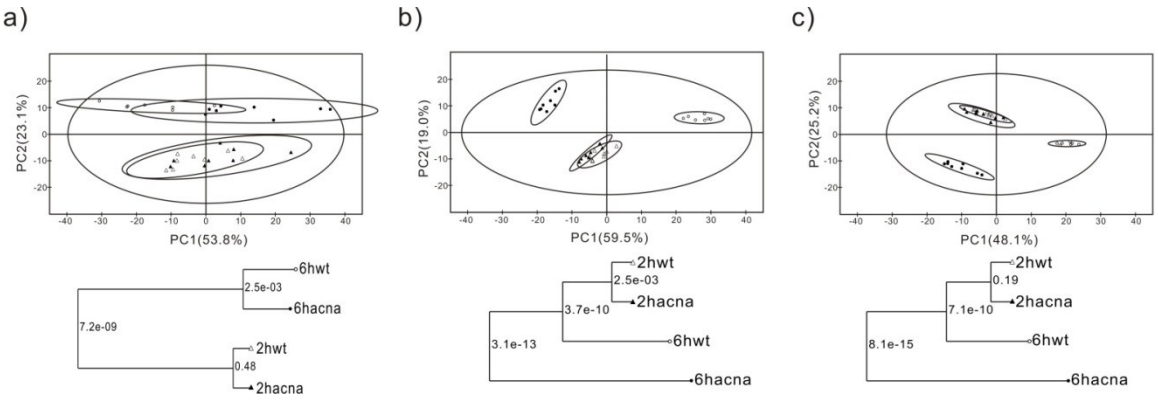


Figure 3.5. Illustration of the impact of NMR preprocessing on within and between class variations in a PCA scores plot. **a)** The 1D ^{1}H NMR spectra was not properly preprocessed. The spectra were not normalized and the noise was not removed. The spectra were only Fourier transformed, phased corrected, and the residual H_2O resonance was removed. **b)** The 1D ^{1}H NMR spectra were processed as in (a) with the addition of normalization using center averaging, but without noise removal. **c)** The 1D ^{1}H NMR spectra were processed as in (b) with the addition of noise removal. Each spectrum was binned using intelligent bucketing with a bin size of 0.025 ppm. The ellipses correspond to the 95% confidence limits from a normal distribution for each cluster. The PCA scores plots compare the metabolomes of *S. aureus* wild-type (wt) strain SA564 with an aconitase mutant (acna) strain SA564-*acnA::tetM* at either two hours (2h) or six hours (6h) of cell growth. Below each PCA scores plot is a corresponding dendrogram generated from the scores using Mahalanobis distances, with *p* values for the null hypothesis reported at each branch.

calculating an average noise value. If a bin across all replicates had an integral value of less than twice the average noise, it was also identified as noise and removed (Figure 3.5b). The protocol for identifying noise regions has been recently improved upon and results in smaller within class variations (Figure 3.5c). This also results in an improved separation between truly distinct classes and removed erroneous separations. For example, the statistical significance between clusters 6hwt and 6hacna improved from a *p*-value of $3.1x10^{-13}$ to $8.1x10^{-15}$, while the small, but biologically irrelevant, separation between clusters 2hwt and 2hacna (*p*-value $2.5x10^{-3}$) was removed (Figure 3.5). Instead of using an average minimal signal intensity to define noise, we now define noise based on a relative standard deviation. This is based on the expectation that real NMR peaks from metabolites will have a higher intrinsic variability compared to the noise because of biological variations that naturally occur even between within class replicates.

Conversely, the variability of the noise should be effectively constant for a given spectrometer operating within normal parameters. Simply, the standard deviation and average is calculated for each bin, where the standard deviation is normalized by theaverage peak intensity. This avoids eliminating weak peaks with a relatively small standard deviation. The same is done for the reference noise region, which is then used to define noise,

$$\textit{Noise}: \quad \sigma_i' \leq \sigma_0' \tag{3.2}$$

$$\textit{Cutoff}: \quad \sigma_0' = \text{avg}(\sigma_n') + 2 * \text{sd}(\sigma_n') \tag{3.3}$$

where σ_i',σ_n' are the relative standard deviations (absolute standard deviation divided by the mean) for the *i*th bin in the spectral region and *n*th bin in the reference noise region,

respectively, and avg(σ_n') and sd(σ_n') are the mean and standard deviation of σ_n' respectively. In effect, any peak that falls within the normal distribution of the reference noise region is defined as a noise bin. This approach is better at defining noise peaks in crowded and overlapping regions of the NMR spectra.

In addition to normalization, each bin or column in the data matrix also needs to be scaled to account for the large dynamic range in peak intensities. PCA and OPLS-DA emphasizes the absolute variation in bins between classes. Correspondingly, the relative variation of an intense peak may be insignificant compared to a weak peak, but the absolute changes in its intensity may completely mask biologically relevant changes in a small peak. Scaling increases the weight of the low intensity peaks so strong peaks do not dominate in PCA and OPLS-DA [78, 79]. In our experience, unit variance scaling, also known as autoscaling or a Z-score (see eqn. 3.1), has been shown to be effective in generating reliable clusters with the correct separation based on biologically relevant class distinctions. Also, within class variance is reduced using autoscaling, which is our default scaling method.

3.5.2 Multivariate statistical analysis of 1D ^{1}H NMR data

We routinely apply PCA, a non-supervised technique, to determine if the 1D ^{1}H NMR data can easily distinguish between the various test classes. PCA provides an unbiased view of group clustering in the resulting 2D scores plot. We only use a three-dimensional (3D) scores plot if class separation in a 2D scores plot is insufficient and the PC_3 contribution is significant (> 5 to 10%). OPLS-DA is only used if class separation is

observed in the PCA scores plot. OPLS-DA is a supervised technique and assesses a relationship between the NMR data class designations. We limit OPLS-DA to only two class designations that differentiate between the single control group (0) and the entire treatment group (1). As a supervised technique, OPLS-DA maximizes a separation between these two classified groups, while minimizing within class variations [39]. Thus, OPLS-DA identifies the important spectral features (metabolites) that primarily contribute to class separation. We routinely use an OPLS-DA S-plot or loading plot to readily identify the key metabolites that contribute to class separation. Since OPLS-DA is a supervised technique and can generate a class separation even for random data [80], it is essential to verify the model [46]. But this is also an advantage of OPLS-DA over PCA; the statistical significance of the model is quantified. We cross-validate OPLS-DA models using a modified leave-one-out method [81, 82] and CV-ANOVA [83]. The modified leave-one-out method provides a quality assessment score (Q^2) and R^2 values, where CV-ANOVA provides a standard *p*-value. The theoretical maximum for Q^2 is 1, where a value of ≥ 0.4 is an empirically acceptable value for biological samples [84], but Q^2 does not have a critical value for inferring significance. It is still possible for an invalid model to produce a large Q^2 value. Similarly, the R^2 values only provide a measure of the fit of the data to the model. But large differences between Q^2 and R^2 ($R^2 >> Q^2$) does suggest an over-fit model. Conversely, a *p*-value $<< 0.05$ from CV-ANOVA provides clear validation of the OPLS-DA model.

In addition to validating the OPLS-DA model, it is also extremely important to verify the statistical significance of the clustering patterns in the PCA and OPLS-DA

scores plot. Is the between group difference larger than the within group variations? One key factor is the number of replicate samples. We have previously shown that increasing the number of replicates improves the statistical significance of cluster separation [85]. This finding is also supported by the increase in *p*-values seen with an increase in within class variations (Figure 3.4). Again, increasing the number of replicates improves the statistical significance of the class separation (lower *p*-value) even when within class variation increases. Correspondingly, we routinely use 10 replicates per group in our metabolomics study to improve the likelihood of observing statistical significant class separations.

It is also important to visually define each group or class within the PCA and OPLS-DA scores plot and to classify the statistical significance of the class separation. We developed a PCA and OPLS-DA utilities software package [66] that draws ellipses or ellipsoids around each group cluster in a scores plot, where the ellipse corresponds to the 95% confidence limits from a normal distribution for each cluster. Visual separation of the ellipses infers a class separation. The same software package is also used to generate a metabolomics tree diagram based on the group clusters in the scores plot [66, 85]. Simply, a centroid from each cluster is used to calculate a Mahalanobis distance between clusters, where dendrograms are then generated from the resulting distance metric. The significance of each node (cluster separation) is determined by using standard bootstrapping techniques and returning a bootstrap number [86], where a value above 50 infers a significant separation; or from Hotelling's T^2 and *F*-distributions that returns a *p*-value, where a number $<< 0.05$ infers a statistically significant separation.

Observing a statistically significant difference in the global metabolome between two or more bacterial samples is typically the first objective of a metabolomics investigation. While this difference may infer some biological significance, the ultimate goal is to identify the underlying metabolites and associated pathways that are the primary contributors to the observed class separation in the PCA and OPLS-DA scores plot. One approach is to generate an S-plot (Figure 3.1) from the resulting OPLS-DA analysis. The S-plot identifies the key bins or ^{1}H chemical shifts that are correlated or anti-correlated with the separation between the two classes in an OPLS-DA scores plot. The ^{1}H chemical shifts can then be compared against a number of online NMR metabolomics databases [87-91] to assign the metabolites. Unfortunately, an unambiguous assignment is rarely possible because of the low chemical shift dispersion and the large number of potential metabolites. Instead, 2D NMR experiments combined with the biological knowledge of the system under investigation are required to improve the accuracy of metabolite identification.

3.5.3 Metabolite Identification

3.5.3.1 Automated peak picking of 2D NMR data

2D $^{1}H,^{13}C$ HSQC and $^{1}H,^{1}H$ TOCSY spectra are commonly used for metabolite identification because of the increase in chemical shift resolution achieved by spreading the information out into two-dimensions. Also, the correlation between ^{1}H chemical shifts for each J-coupled H pair; and the correlation between ^{1}H and ^{13}C chemical shifts for each C-H pair significantly reduces the assignment ambiguity. This occurs because

now both chemical shifts have to match a single metabolite in the database to make an assignment. Despite the advantages, peak picking and organizing a table of intensities from a 2D NMR experiment is a time consuming process, especially when multiple spectra are involved. Numerous software packages are available to automate the peak picking of 2D NMR spectra, however; it is extremely difficult, if not impossible, to align and match multiple sets of spectra with different peak patterns due to unique metabolomes.

For example, three different sets of cell cultures (different cell types, treatments or environmental conditions, etc.) will each exhibit a distinct set of peaks in the NMR spectrum due to the presence of unique metabolites. These unique peaks will be mixed with other peaks common to all three groups, but the relative peak intensities are likely to vary due to different metabolite concentrations. Thus, if the control group is designated as the reference spectrum for automated peak picking, a peak list will be generated that only contains peaks observed in the control spectrum that are above the designated noise threshold. Correspondingly, peaks unique to the other two groups will be missed when this peak list is used to peak pick their spectrum. In addition, weak peaks may also be missed due to different noise levels between the spectra and a corresponding difference in the threshold setting for peak picking. Instead, a composite reference spectrum for automated peak picking needs to be generated that captures *all* the peaks present in the three separate groups. We accomplish this task by using the addNMR function in the NMRpipe software package [92]. As the name implies, addNMR mathematically sums all spectra together from the three groups to make a single spectrum. This resulting

"master spectrum" contains all the peaks observed throughout the set of 2D experiments and is used to generate a peak list for automated peak picking of each individual spectrum. Critically, the 2D NMR spectra need to be collected and processed using identical experimental parameters (spectral width, data points, zero-filling, etc.) and needs to be aligned to an internal reference (TMSP-d_4). In our experience, all the peaks from the complete set of NMR spectra are routinely matched to the reference list by using a chemical shift error-tolerance of 0.04 ppm and 0.25 ppm in the ^{1}H and ^{13}C dimensions, respectively. This approach has greatly simplified and increased the efficiency of a previously laborious procedure. The addNMR command can also be used to generate a difference spectrum that clearly highlights the major spectral changes between two classes (Figure 3.6). cultures. Positive peaks, increased metabolite concentration, are colored green and negative peaks, decreased metabolite concentration are colored red.

3.5.3.2 Assignment of an NMR peak to a metabolite

Metabolite identification is an extremely important component of the metabolomics process because it enables the determination of the key metabolites perturbed by the treatment or the metabolites primarily contributing to class distinction. This includes the discovery of important biomarkers associated with drug efficacy or drug resistance. Also, metabolite identification is important to the drug discovery process by either identifying metabolic pathways affected by a drug to evaluate efficacy or potential toxicity; or by identifying potentially new therapeutic targets. Nevertheless, accurate metabolite identification is very difficult and labor-intensive. The success of

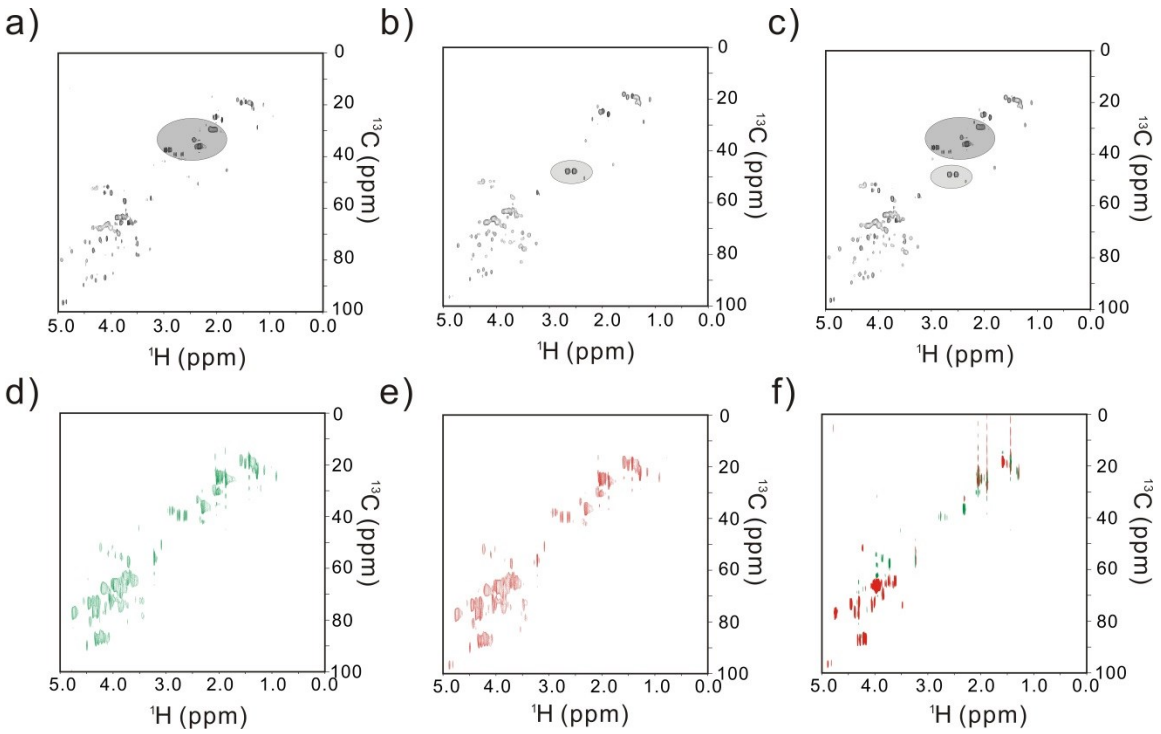


Figure 3.6. a-c) Illustration of the procedure to generate a "master spectrum" and facilitate automated peak picking by creating a complete peak list. (a-b) Representative 2D ^{1}H-^{13}C HSQC spectra obtained from two distinct bacterial cultures, where some major spectral differences are highlighted. **c)** The two 2D ^{1}H-^{13}C HSQC spectra from (a-b) were added to yield a master spectrum that contains all the observed NMR peaks. **d-f)** Illustration of the procedure to generate a "difference spectrum" to facilitate metabolite identification by creating a signed (+, -, null) peak list. (d-e) Representative 2D ^{1}H-^{13}C HSQC spectra obtained from two distinct bacterial cultures. **f)** The two 2D ^{1}H-^{13}C HSQC spectra from (d-e) were subtracted to yield a difference spectrum that identifies the NMR peaks, and correspondingly metabolites, that differ between the two bacterial cell

metabolite spectral assignment relies largely on the completeness of metabolomics databases. We routinely use a combination of the following databases: Human Metabolome Database [87], Madison Metabolomics Consortium Database [88], Platform for RIKEN Metabolomics [89], BioMagResBank [90], and Metabominer [91], which provide both redundant and complementary NMR spectral data. Importantly, the reference NMR spectra in the various databases were obtained under different buffer condition and use different internal standards. This results in a range of potential chemical shifts for a given metabolite. Thus, the databases with sample conditions that closely match our experimental conditions are used for chemical shift matching. The overall goal is to identify a complete set of metabolites as quickly and accurately as possible without any bias, by matching the experimental chemical shifts from the 2D NMR spectra with the values in the database.

For a 2D ^{1}H-^{13}C HSQC experiment, it is important to realize that metabolites may be heterogeneously labeled by the carbon-13 source present in the growth media. Correspondingly, all the peaks for a specific metabolite may not be detectable in the 2D ^{1}H-^{13}C HSQC experiment. Also, a reference spectrum for the metabolite may not be present in any of the available databases. The assignment of a particular peak might still be ambiguous because multiple metabolites may contain the same chemical shift or contain an identical substructure (i.e., ATP, ADP, AMP or NAD, NADPH). Therefore, a few automated filters are applied to overcome some of these ambiguities during the peak assignment process.

The first filter is to verify that the bacteria can actually produce the proposed metabolite. This is routinely accomplished by searching the Biocyc [93] and KEGG [94] database for metabolites known to exist in the bacteria under investigation. The second filter is based on a differential peak list. All the NMR peaks potentially assigned to a specific metabolite should have the same trend relative to the control. Obviously, the metabolite can only have one concentration and all the NMR peaks need to be consistent with this single concentration. Correspondingly, all the peaks have to be increased, decreased or the same relative to the same peaks in the control spectrum. This is easily and quickly visualized by subtracting the two sets of spectra and generating a signed (+, -, null) peak list. Peaks assigned to the same metabolite have to have the same sign. The third filter is based on a biological relationship with other metabolites. Simply, the likelihood of a correct assignment increases if other metabolites in a specific metabolic pathway have also been assigned. It is more likely to observe multiple metabolites from the same pathway than various metabolites from unrelated pathways. Similarly, if there is a direct metabolic path between two or more metabolites, then their assignments are more likely to be correct. The final filter is the application of our biological knowledge of the bacterial system under investigation. The pathways or metabolites that are expected to be perturbed by the treatment would be given precedent in the assignment process. As a simple example, a comparison between wild-type and mutant bacterial strains where aconitase has been inactivated would reasonably be expected to lead to changes in metabolites associate with the TCA cycle. Likewise, a comparison between untreated and

drug-treated cells would be expected to lead to changes in metabolic pathways inhibited by the drug.

3.5.3.3 Statistical analysis of the 2D ^{1}H-^{13}C HSQC data.

After assigning the 2D ^{1}H-^{13}C HSQC spectra to a set of metabolites, the next goal is to determine metabolite concentration differences between the various bacterial culture conditions under investigation. Unfortunately, peak intensities in a standard 2D ^{1}H-^{13}C HSQC experiment are dependent on metabolite concentrations and J-couplings, dynamics, and relaxation properties [76]. Therefore, only a relative percent change in a metabolite concentration can be determined [5]. Alternatively, an absolute concentration can be determined using $HSQC_0$, which requires a set of three HSQC experiments per sample. We routinely employ both approaches [76].

A relative difference in peak intensities is determined by using a triplicate set of a conventional 2D ^{1}H-^{13}C HSQC experiment for each bacterial culture condition. Prior to calculating a relative percent change in peak intensities, a detailed normalization process is required, which was previously described in detail [5]. First, the peak intensities within each spectrum are normalized by dividing each peak by the internal standard, the intensity of the TMSP-d_4 peak. Each peak pertaining to a specific chemical shift across each triplicate data set is then normalized by the most intense peak in the set of three peaks. Specifically, the maximal intensity for each peak across all data sets would be set to 100 and all other intensities are scaled relative to this peak intensity. Then all the normalized intensity for a given metabolite for each triplicate set is averaged together,

and a relative percent error can be calculated between different cultures. A Student's t-test or ANOVA is then used to determine if the relative change in peak intensities is statistically significant at the 95% confidence limit. Calculating a relative difference in metabolite concentrations can be beneficial to understanding broader changes to the system, especially when a cluster of metabolites in a specific pathway exhibit a similar trend in concentration changes inferring an important role for the metabolic pathway. Nevertheless, this approach is rather cumbersome and does not allow for a direct comparison between different metabolites.

Alternatively, we routinely use the $HSQC_0$ experiment to determine absolute metabolite concentrations. The overall protocol for the extrapolation of peak intensities to time-zero and the determination of the associated concentration has been previously described in detail [76, 77]. As an example, Figure 3.7 shows a linear regression plot of the peak intensities for each carbon-hydrogen pair in fructose. This demonstrates that the average peak height can be extrapolated. A distinct advantage of this method is that a single calibration curve can be made using multiple compounds with known concentrations to correlate the time-zero peak intensity with a concentration. Figure 3.8 illustrates such a calibration curve using 5 different mixtures, each consisting of 9 different ^{13}C-labeled metabolites ranging in concentrations from 5 to 300 μM. Also, the concentration for each metabolite was randomized in each mixture. For example the concentration of ^{13}C-D-alanine in the 5 mixtures is 300, 10, 25, 5, and 100 μM, respectively. The data was fitted using a weighted linear least squares calculation. Notably, the best-fit line (R^2 0.997) has a y-intercept close to zero as expected for a

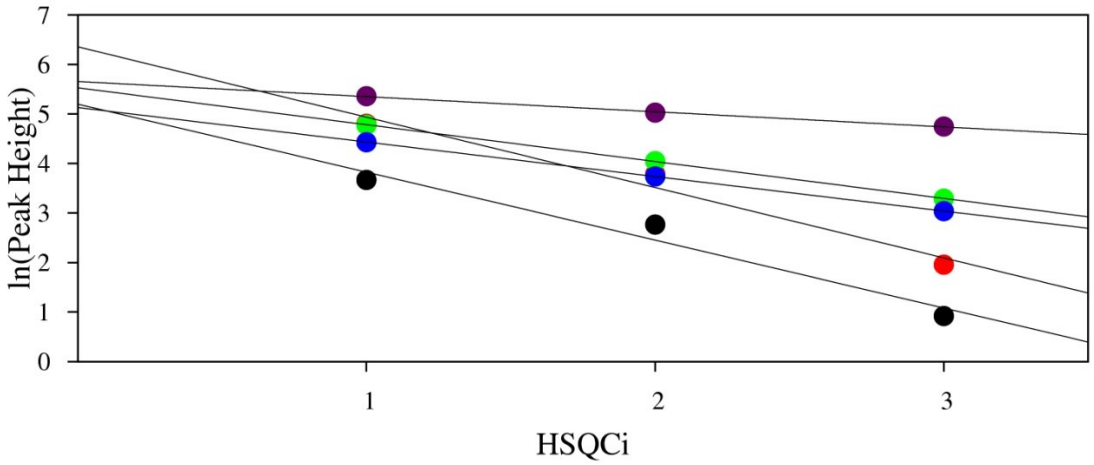


Figure 3.7. Extrapolation of 2D HSQCi (i=1,2,3) peak intensities of fructose. Fructose consists of multiple peaks in the 2D spectrum where each symbol represents a different peak. The attenuation factors (spin dynamic properties) for each C-H pair represent the slope of the line. The lines converge to the same general region at the y-intercept, where the intensity is directly related to concentration.

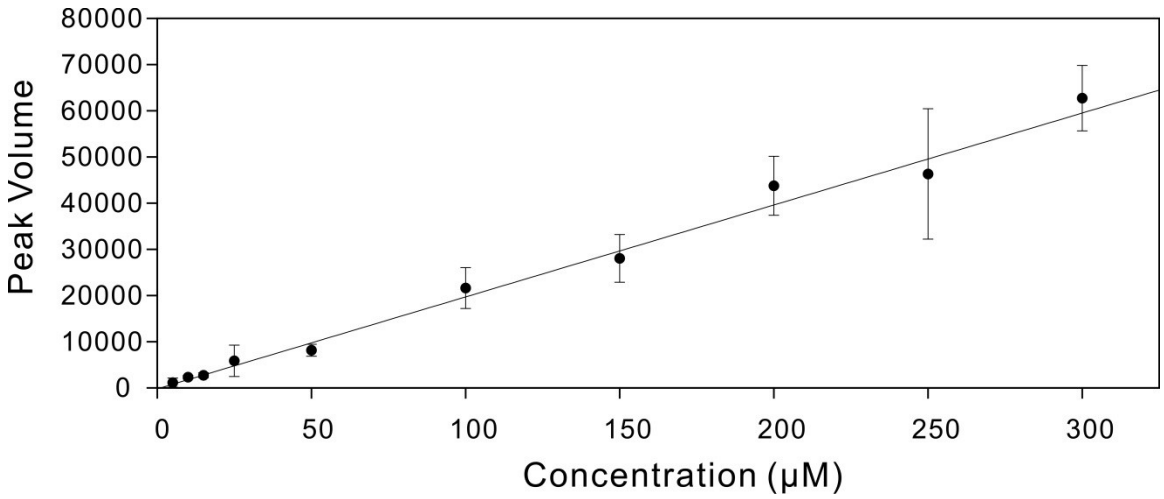


Figure 3.8. A strong correlation between NMR peak volumes and metabolite concentrations (R^2 0.997) is demonstrated by linear regression plot generated from $HSQC_0$ data. $HSQC_0$ NMR spectra were collected for five different metabolite mixtures containing nine ^{13}C-labeled compounds with concentrations ranging from 5 μM to 300 μM. The relationship between peak volume and metabolite concentration is independent of the metabolite.

concentration of zero. Also, the correlation between peak volume and concentration is independent of the metabolite. Importantly, the accurate application of the calibration curve requires collecting and processing $HSQC_0$ spectra *identical* to the parameters used to obtain the original calibration curve. Critically, the receiver gain must be the same for all samples, because any change in the receiver gain influences the slope of the calibration curve. Also, the addition of 500 μM TMSP-d_4 as an internal standard is crucial, because both the calibration samples and experimental samples must both be similarly normalized to the TMSP-d4 peak. As an example, if the TMSP-d4 peak volumes for the calibration mixtures are 1000, 500, and 250 for each $HSQC_i$ (i = 1,2,3) spectrum, respectively, then the experimental results for all *in vivo* metabolite extracts must be normalized so that the internal standard (TMSP-d_4) peak volumes are also 1000, 500, and 250. The concentrations are measured in triplicate, where a Student's t-test or ANOVA is used to determine if the concentration changes are statistically significant at the 95% confidence limit.

3.5.4 Metabolomics Network Map

Metabolites are highly interconnected through numerous metabolic pathways that form an extremely complex network [95]. Correspondingly, it is not uncommon to observe correlated changes between distantly connected metabolites. In effect, metabolomics depends on these complex interactions to understand the phenotype of a bacterial cell. Thus, a metabolomics network map provides an efficient approach to visualize and summarize the overall changes to the metabolome, to validate metabolite

assignments based on clear connections to other metabolites, and the identification of key metabolic pathways.

We have routinely used Cytoscape to easily and quickly generate metabolomics network maps. Cytoscape is a user-friendly software package with plug-ins related to metabolomics [96-99]. Cytoscape simply requires a list of the metabolites and their associated concentration changes as input. The connections between nodes (metabolites) in the map are based on metabolic pathways from the MetaCyc database [100]. An example of a typical Cytoscape map summarizing the observed changes in the *S. epidermidis* metabolome caused by environmental stimuli associated with biofilm formation is shown in Figure 3.1. The metabolomics network map can be easily modified to highlight specific features of the metabolome. Edges can be broadened to highlight specific pathways; and the color and size of nodes can be adjusted to reflect the direction and magnitude of the concentration changes, respectively [101]. Cytoscape also provides a range of map design choices. Unfortunately, the resulting network maps (Figure 3.1) do not resemble standard metabolic pathways. Thus, Cytoscape maps are simply used as a template to manually draw more traditional looking metabolic pathways. Since Cytoscape maps are so easily generated, we also use the software to assist in metabolite assignments. Potential lists of metabolite assignments are input into Cytoscape to identify metabolites that are isolated nodes excluded from the main network map. These metabolites are likely missed assigned and are reevaluated. In addition to Cytoscape, we also use the R statistics package to create heat maps from absolute metabolite concentrations or percent relative concentration changes.

3.6 Conclusion

NMR metabolomics is an invaluable tool for systems biology and its application is rapidly expanding. Global changes in the metabolic state of bacterial cells occur as a result of environmental stressors, genetic modifications, drug treatments, or numerous other factors. A detailed analysis of the differences in the NMR spectra is commonly used to identify the key metabolite changes that differentiate between these bacterial classes (i.e., controls versus treated). In addition, metabolite identification by NMR allows for the subsequent identification of the important metabolic pathways that are affected by the treatment, providing further insight into the underlying biological process. The appeal of NMR metabolomics is its simplicity, but unfortunately it is also easy to obtain unreliable results. The observed changes in the metabolome should be biologically relevant, but because the metabolome is so sensitive to any environmental change; it is also easily perturbed by the experimental protocol. This is clearly an undesirable outcome. To address this issue, we described in detail our optimized protocols for the NMR analysis of bacterial metabolomes. We also highlighted common problems and potential sources of mistakes. We discuss the entire process that includes growing and harvesting bacterial cells, extracting the metabolome, NMR data collection, processing and analysis, statistical analysis, metabolite and network identification. The protocols described have been successfully applied to a number of systems biology projects.

3.7 References

1. Shyur, L.F. and N.S. Yang, *Metabolomics for Phytomedicine Research and Drug Development.* Current Opinion in Chemical Biology, 2008. **12**(1): p. 66-71.

2. Spratlin, J.L., N.J. Serkova, and S.G. Eckhardt, *Clinical Applications of Metabolomics in Oncology: A Review.* Clinical Cancer Research, 2009. **15**(2): p. 431-40.

3. Rochfort, S., *Metabolomics Reviewed: A New "Omics" Platform Technology for Systems Biology and Implications for Natural Products Research.* Journal of Natural Products, 2005. **68**(12): p. 1813-20.

4. Raamsdonk, L.M., et al., *A Functional Genomics Strategy That Uses Metabolome Data to Reveal the Phenotype of Silent Mutations.* Nature Biotechnology, 2001. **19**(1): p. 45-50.

5. Zhang, B., et al., *NMR Analysis of a Stress Response Metabolic Signaling Network.* Journal Proteome Research, 2011. **10**(8): p. 3743-54.

6. Ewald, J.C., T. Matt, and N. Zamboni, *The Integrated Response of Primary Metabolites to Gene Deletions and the Environment.* Molecular BioSystems, 2013. **9**(3): p. 440-6.

7. Fiehn, O., *Metabolomics--The Link Between Genotypes and Phenotypes.* Plant Molecular Biology, 2002. **48**(1-2): p. 155-71.

8. Kaddurah-Daouk, R., B.S. Kristal, and R.M. Weinshilboum, *Metabolomics: A Global Biochemical Approach to Drug Response and Disease.* Annual Review of Pharmacology and Toxicology, 2008. **48**: p. 653-83.

9. Keun, H.C., *Metabonomic Modeling of Drug Toxicity.* Pharmacology & Therapeutics, 2006. **109**(1-2): p. 92-106.

10. Zhang, W., F. Li, and L. Nie, *Integrating Multiple 'Omics' Analysis for Microbial Biology: Application and Methodologies.* Microbiology, 2010. **156**(Pt 2): p. 287-301.

11. Ideker, T., et al., *Integrated Genomic and Proteomic Analyses of a Systematically Perturbed Metabolic Network.* Science, 2001. **292**(5518): p. 929-34.

12. Pir, P., et al., *Integrative Investigation of Metabolic and Transcriptomic Data.* BMC Bioinformatics, 2006. **7**: p. 203.

13. Mosier, A.C., et al., *Metabolites Associated with Adaptation of Microorganisms to an Acidophilic, Metal-Rich Environment Identified by Stable-Isotope-Enabled Metabolomics.* mBio, 2013. **4**(2).

14. Sadykov, M.R., et al., *Using NMR Metabolomics to Investigate Tricarboxylic Acid Cycle-Dependent Signal Transduction in Staphylococcus epidermidis.* Journal of Biological Chemistry, 2010. **285**(47): p. 36616-36624.

15. Yoon, S.H., et al., *Comparative Multi-Omics Systems Analysis of Escherichia coli Strains B and K-12.* Genome Biology, 2012. **13**(5): p. R37.

16. Yang, R., et al., *Omics Strategies for Revealing Yersinia pestis Virulence.* Frontiers in Cellular and Infection Microbiology, 2012. **2**: p. 157.

17. Bisson, G.P., et al., *Upregulation of the Phthiocerol dimycocerosate Biosynthetic Pathway by Rifampin-Resistant, rpoB Mutant Mycobacterium tuberculosis.* Journal of Bacteriology, 2012. **194**(23): p. 6441-52.

18. Nobeli, I., et al., *A Structure-Based Anatomy of the E. coli Metabolome.* J. Mol. Biol., 2003. **334**(Copyright (C) 2013 American Chemical Society (ACS). All Rights Reserved.): p. 697-719.

19. Weckwerth, W., *Metabolomics: An Integral Technique in Systems Biology.* Bioanalysis, 2010. **2** American Chemical Society, 2010. **2**: p. 829-836.

20. Powers, R., *NMR Metabolomics and Drug Discovery.* Magnetic Resonance in Chemistry, 2009. **47 Suppl 1**: p. S2-11.

21. Moco, S., et al., *Metabolomics Technologies and Metabolite Identification.* Trends in Analytical Chemistry, 2007. **26**(9): p. 855-866.

22. Villas-Boas, S.G., et al., *Mass Spectrometry in Metabolome Analysis.* Mass Spectrometry Review, 2005. **24** p. 613-646.

23. Wilson, I.D., et al., *HPLC-MS-Based Methods for the Study of Metabonomics.* Journal Chromatography, 2005. **817**: p. 67-76.

24. Reo, N.V., *NMR-Based Metabolomics.* Drug Chemical Toxicology, 2002. **25**: p. 375-382.

25. Fan, T.W.M. and A.N. Lane, *NMR-Based Stable Isotope Resolved Metabolomics in Systems Biochemistry.* Journal Biomolecular NMR, 2011. **49**): p. 267-280.

26. Lewis, I.A., et al., *NMR Method for Measuring Carbon-13 Isotopic Enrichment of Metabolites in Complex Solutions.* Analytical Chemistry, 2010. **82**: p. 4558-4563.

27. Kell, D.B., *Metabolomics and Systems Biology: Making Sense of the Soup.* Current Opinion Microbiology, 2004. **7**: p. 296-307.

28. Metz, T.O., et al., *High-Resolution Separations and Improved Ion Production and Transmission in Metabolomics.*Trends Analytical Chemistry, 2008. **27**: p. 205-214.

29. Pan, Z. and D. Raftery, *Comparing and Combining NMR Spectroscopy and Mass Spectrometry in Metabolomics.* Analytical Bioanalytical Chemistry, 2007. **387**: p. 525-527.

30. Viant, M.R., E.S. Rosenblum, and R.S. Tjeerdema, *NMR-Based Metabolomics: A Powerful Approach for Characterizing the Effects of Environmental Stressors on Organism Health.* Environmental Science & Technology, 2003. **37**(21): p. 4982-4989.

31. Beckonert, O., et al., *Metabolic Profiling, Metabolomic and Metabonomic Procedures for NMR Spectroscopy of Urine, Plasma, Serum and Tissue Extracts.* Nature Protocol, 2007. **2**(11): p. 2692-2703.

32. Dietmair, S., et al., *Towards Quantitative Metabolomics of Mammalian Cells: Development of a Metabolite Extraction Protocol.* Analytical Biochemistry, 2010. **404**(2): p. 155-164.

33. Kim, H.K., Y.H. Choi, and R. Verpoorte, *NMR-Based Metabolomic Analysis of Plants.* Nature Protocols, 2010. **5**(3): p. 536-49.

34. Marcinowska, R., et al., *Optimization of a Sample Preparation Method for the Metabolomic Analysis of Clinically Relevant Bacteria.* Journal Microbiological Methods, 2011. **87**(1): p. 24-31.

35. Mashego, M., et al., *Microbial Metabolomics: Past, Present and Future Methodologies.* Biotechnology Letters, 2007. **29**(1): p. 1-16.

36. Wu, X.H., et al., *Sampling Methods for NMR-Based Metabolomics of Staphylococcus aureus.* Journal of Biotechnology, 2010. **5**(1): p. 75-84.

37. Cuperlovic-Culf, M., et al., *Cell Culture Metabolomics: Applications and Future Directions.* Drug Discovery Today, 2010. **15**(15-16): p. 610-21.

38. Ramadan, Z., et al., *Metabolic Profiling using Principal Component Analysis, Discriminant Partial Least Squares, and Genetic Algorithms.* Talanta, 2006. **68**(5): p. 1683-91.

39. Bylesjo, M., et al., *OPLS Discriminant Analysis: Combining the Strengths of PLS-DA and SIMCA Classification.* Journal of Chemometrics, 2006. **20**(8-10): p. 341-351.

40. Fernie, A.R., et al., *Innovation - Metabolite Profiling: From Diagnostics to Systems Biology.* Nature Reviews Molecular Cell Biology, 2004. **5**(9): p. 763-769.

41. Clarke, C.J. and J.N. Haselden, *Metabolic Profiling as a Tool for Understanding Mechanisms of Toxicity.* Toxicologic Pathology, 2008. **36**(1): p. 140-147.

42. Faijes, M., A.E. Mars, and E.J. Smid, *Comparison of Quenching and Extraction Methodologies for Metabolome Analysis of Lactobacillus plantarum.* Microbial Cell Factories, 2007. **6**: p. 27.

43. Van Batenburg, M.F., et al., *New Figures of Merit for Comprehensive Functional Genomics Data: The Metabolomics Case.* Analytical Chemistry. , 2011. **83**: p. 3267-3274.

44. Kanani, H., P.K. Chrysanthopoulos, and M.I. Klapa, *Standardizing GC-MS Metabolomics.* Journal Chromatography B, 2008. **871**(2): p. 191-201.

45. Teahan, O., et al., *Impact of Analytical Bias in Metabonomic Studies of Human Blood Serum and Plasma.* Analytical Chemistry, 2006. **78**: p. 4307-4318.

46. Worley, B. and R. Powers, *Multivariate Analysis in Metabolomics.* Current Metabolomics, 2013. **1**(1): p. 92-107.

47. Birkemeyer, C., et al., *Metabolome Analysis: The Potential of in vivo Labeling with Stable Isotopes for Metabolite Profiling.* Trends Biotechnology, 2005. **23**(1): p. 28-33.

48. Fan, T.M., et al., *Metabolomics-Edited Transcriptomics Analysis of Se Anticancer Action in Human Lung Cancer Cells.* Metabolomics, 2005. **1**(4): p. 325-339.

49. Halouska, S., et al., *Predicting the in Vivo Mechanism of Action for Drug Leads Using NMR Metabolomics.* ACS chemical biology, 2012. **7**(1): p. 166-171.

50. Halouska, S., et al., *Use of NMR Metabolomics to Analyze the Targets of D-Cycloserine in Mycobacteria: Role of D-Alanine Racemase.* Journal of Proteome Research, 2007. **6**(12): p. 4608-14.

51. Forgue, P., et al., *NMR Metabolic Profiling of Aspergillus nidulans to Monitor Drug and Protein Activity.* Journal of Proteome Research, 2006. **5**(8): p. 1916-1923.

52. Suhre, K. and C. Gieger, *Genetic Variation in Metabolic Phenotypes: Study Designs and Applications.* Nature Reviews Genetics, 2012. **13**(11): p. 759-769.

53. Halouska, S. and R. Powers, *Negative Impact of Noise on the Principal Component Analysis of NMR Data.* Journal of Magnetic Resonance, 2006. **178**(1): p. 88-95.

54. Bolten, C.J., et al., *Sampling for Metabolome Analysis of Microorganisms.* Analytical Chemistry, 2007. **79**(10): p. 3843-9.

55. Bailey, N.J.C., et al., *An NMR-based Metabolomic Approach to the Analysis of the Effects of Xenobiotics on Endogenous Metabolite Levels in Plants.* Spectroscopy-An International Journal, 2004. **18**(2): p. 279-287.

56. Defernez, M. and I.J. Colquhoun, *Factors Affecting the Robustness of Metabolite Fingerprinting using 1H NMR Spectra.* Phytochemistry, 2003. **62**(6): p. 1009-17.

57. Broadhurst, D.I. and D.B. Kell, *Statistical Strategies for Avoiding False Discoveries in Metabolomics and Related Experiments.* Metabolomics, 2006. **2**(4): p. 171-196.

58. Brereton, R.G., *Consequences of Sample Size, Variable Selection, and Model Validation and Optimisation, for Predicting Classification Ability from Analytical Data.* Trends in Analytical Chemistry, 2006. **25**(11): p. 1103-1111.

59. Canelas, A., et al., *Leakage-Free Rapid Quenching Technique for Yeast Metabolomics.* Metabolomics, 2008. **4**(3): p. 226-239.

60. Saude, E.J. and B.D. Sykes, *Urine Stability for Metabolomic Studies: Effects of Preparation and Storage.* Metabolomics, 2007. **3**: p. 19-27.

61. Giuliodori, A.M., et al., *Review on Bacterial Stress Topics.* Ann. N. Y. Academy of Science, 2007. **1113**: p. 95-104.

62. Schadel, F., F. David, and E. Franco-Lara, *Evaluation of Cell Damage Caused by Cold Sampling and Quenching for Metabolome Analysis.* Applied Microbiology and Biotechnology, 2011. **92**(6): p. 1261-1274.

63. Wellerdiek, M., et al., *Metabolic Quenching of Corynebacterium glutamicum: Efficiency of Methods and Impact of Cold Shock.* Bioprocess and Biosystems Engineering, 2009. **32**(5): p. 581-592.

64. Wittmann, C., et al., *Impact of the Cold Shock Phenomenon on Quantification of Intracellular Metabolites in Bacteria.* Analytical Biochemistry, 2004. **327**(1): p. 135-139.

65. Mazo, R.M., *A Fluctuation Theory Analysis of the Salting-Out Effect.* Journal of Physical Chemistry, 2006. **110**: p. 24077-24082.

66. Worley, B., S. Halouska, and R. Powers, *Utilities for Quantifying Separation in PCA/PLSA-DA Scores.* Analytical Biochemistry, 2013. **433**(2): p. 102-104.

67. Potts, B.C.M., et al., *NMR of Biofluids and Pattern Recognition: Assessing the Impact of NMR Parameters on the Principal Component Analysis of Urine from Rat and Mouse.* Journal of Pharmacy Biomedical Analysis, 2001. **26**: p. 463-476.

68. Hoult, D.I., *Solvent Peak Saturation with Single-Phase and Quadrature Fourier Transformation.* Journal of Magnetic Resonance, 1976. **21**(2): p. 337-347.

69. Sklenar, V., et al., *Gradient-Tailored Water Suppression for H-1-N-15 Hsqc Experiments Optimized to Retain Full Sensitivity.* Journal of Magnetic Resonance Series A, 1993. **102**(2): p. 241-245.

70. Liu, M.L., et al., *Improved WATERGATE Pulse Sequences for Solvent Suppression in NMR Spectroscopy.* Journal of Magnetic Resonance, 1998. **132**(1): p. 125-129.

71. Ogg, R.J., P.B. Kingsley, and J.S. Taylor, *Wet, a T-1-Insensitive and B-1-Insensitive Water-Suppression Method for in-Vivo Localized H-1-Nmr Spectroscopy.* Journal of Magnetic Resonance Series B, 1994. **104**(1): p. 1-10.

72. Simpson, A.J. and S.A. Brown, *Purge NMR: Effective and Easy Solvent Suppression.* Journal of Magnetic Resonance, 2005. **175**(2): p. 340-346.

73. Hwang, T.L. and A.J. Shaka, *Water Suppression That Works - Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients.* Journal of Magnetic Resonance Series A, 1995. **112**(2): p. 275-279.

74. McKenzie, J., et al., *Peak fitting in 2D 1H–13C HSQC NMR Spectra for Metabolomic Studies.* Metabolomics, 2010. **6**(4): p. 574-582.

75. Xi, Y., et al., *Improved Identification of Metabolites in Complex Mixtures using HSQC NMR Spectroscopy.* Analytica Chimica Acta, 2008. **614**(2): p. 127-33.

76. Hu, K.F., W.M. Westler, and J.L. Markley, *Simultaneous Quantification and Identification of Individual Chemicals in Metabolite Mixtures by Two-Dimensional Extrapolated Time-Zero H-1-C-13 HSQC (HSQC(0)).* Journal of the American Chemical Society, 2011. **133**(6): p. 1662-1665.

77. Hu, K.F., et al., *Measurement of Absolute Concentrations of Individual Compounds in Metabolite Mixtures by Gradient-Selective Time-Zero H-1-C-13 HSQC with Two Concentration References and Fast Maximum Likelihood Reconstruction Analysis.* Analytical Chemistry, 2011. **83**(24): p. 9352-9360.

78. van den Berg, R.A., et al., *Centering, Scaling, and Transformations: Improving the Biological Information Content of Metabolomics Data.* BMC Genomics, 2006. **7**: p. 142.

79. Craig, A., et al., *Scaling and Normalization Effects in NMR Spectroscopic Metabonomic Data Sets.* Analytical Chemistry, 2006. **78**(7): p. 2262-2267.

80. Kjeldahl, K. and R. Bro, *Some Common Misunderstandings in Chemometrics.* Journal of Chemometrics, 2010. **24**(7-8): p. 558-564.

81. Shao, J., *Linear-Model Selection by Cross-Validation.* Journal of the American Statistical Association, 1993. **88**(422): p. 486-494.

82. Golbraikh, A. and A. Tropsha, *Beware of q(2)!* Journal of Molecular Graphics & Modelling, 2002. **20**(4): p. 269-276.

83. Eriksson, L., J. Trygg, and S. Wold, *CV-ANOVA for Significance Testing of PLS and OPLS® models.* Journal of Chemometrics, 2008. **22**(11-12): p. 594-600.

84. Westerhuis, J., et al., *Assessment of PLSDA Cross Validation.* Metabolomics, 2008. **4**(1): p. 81-89.

85. Werth, M.T., et al., *Analysis of Metabolomic PCA Data using Tree Diagrams.* Analytical Biochemistry, 2010. **399**(1): p. 58-63.

86. Henderson, A.R., *The Bootstrap: A Technique for Data-Driven Statistics. Using Computer-Intensive Analyses to Explore Experimental Data.* Clinical Chimica Acta, 2005. **359**: p. 1-26.

87. Wishart, D.S., et al., *HMDB: A Knowledgebase for the Human Metabolome.* Nucleic Acids Research, 2009. **37**: p. D603-D610.

88. Cui, Q., et al., *Metabolite Identification via the Madison Metabolomics Consortium Database.* Nature Biotechnology, 2008. **26**(2): p. 162-4.

89. Akiyama, K., et al., *PRIMe: a Web Site that Assembles Tools for Metabolomics and Transcriptomics.* In Silico Biology, 2008. **8**(3-4): p. 339-45.

90. Ulrich, E.L., et al., *BioMagResBank.* Nucleic Acids Research, 2008. **36**(Database issue): p. D402-8.

91. Xia, J., et al., *MetaboMiner--Semi-Automated Identification of Metabolites from 2D NMR Spectra of Complex Biofluids.* BMC Bioinformatics, 2008. **9**: p. 507.

92. Delaglio, F., et al., *NMRPipe: A Multidimensional Spectral Processing System Based on UNIX pipes.* Journal of Biomolecular NMR, 1995. **6**(3): p. 277-93.

93. Caspi, R., et al., *The MetaCyc Database of Metabolic Pathways and Enzymes and the BioCyc Collection of Pathway/Genome Databases.* Nucleic Acids Research, 2012. **40**(Database issue): p. D742-53.

94. Kanehisa, M., et al., *KEGG for Linking Genomes to Life and the Environment.* Nucleic Acids Research, 2008. **36**: p. D480-D484.

95. Kohlstedt, M., J. Becker, and C. Wittmann, *Metabolic Fluxes and Beyond—Systems Biology Understanding and Engineering of Microbial Metabolism.* Applied Microbiology and Biotechnology, 2010. **88**(5): p. 1065-1075.

96. Bot, J.J. and M.J. Reinders, *CytoscapeRPC: A Plugin to Create, Modify and Query Cytoscape Networks From Scripting Languages.* Bioinformatics, 2011. **27**(17): p. 2451-2.

97. Gao, J., et al., *Metscape: a Cytoscape Plug-in for Visualizing and Interpreting Metabolomic Data in the Context of Human Metabolic Networks.* Bioinformatics, 2010. **26**(7): p. 971-3.

98. Kohl, M., S. Wiese, and B. Warscheid, *Cytoscape: Software for Visualization and Analysis of Biological Networks.* Methods in Molecular Biology, 2011. **696**: p. 291-303.

99. Smoot, M.E., et al., *Cytoscape 2.8: New Features for Data Integration and Network Visualization.* Bioinformatics, 2011. **27**(3): p. 431-2.

100. Caspi, R., et al., *The MetaCyc Database of Metabolic Pathways and Enzymes and the BioCyc Collection of Pathway/Genome Databases.* Nucleic Acids Research, 2008. **36**: p. D623-D631.

101. Chaika, N.V., et al., *MUC1 Mucin Stabilizes and Activates Hypoxia-Inducible Factor 1 Alpha to Regulate Metabolism in Pancreatic Cancer.* Proceeding for the National Academy of Science U S A, 2012. **109**(34): p. 13787-92.

CHAPTER 4

PREDICTING THE *IN VIVO* MECHANISM OF ACTION FOR DRUG LEADS USING NMR METABOLOMICS[§]

4.1 Introduction

Emerging and remerging infectious disease outbreaks from numerous gram-negative and gram positive pathogens have increased dramatically over the past decade [1]. Further, we are facing the serious likelihood that these pathogens will soon become resistant to all known antibacterial treatments, which may lead to worldwide pandemics [2]. Unfortunately, the development and approval of antibiotics have not kept pace with the growing emergence of resistant pathogens [3]. Instead, there has been a decline in the approval of new antibiotics [4]. Twenty novel classes of marketable antibiotics were produced between 1930 and 1962 [5]. These classes of antibiotics inhibit a short list of cellular processes: cell wall biosynthesis, DNA supercoiling, transcription, translation and folate biosynthesis. Since 1962, only two new antibiotic classes have received FDA approval: oxazolidinones, which inhibits protein synthesis, and cyclic lipopeptides, which destroys membrane potential. Both compounds are used in the treatment of gram positive bacteria, such as methicillin-resistant *Staphylococcus aureus* (MRSA) [5]. However, additional antibiotics are needed to combat the prevalence of other multidrug resistant pathogens, such as *Enterococcus faecium*, *Klebsiella pneumonia*, *Acinetobacter*

[§] Chapter 4 was adapted from Halouska S., *et. al*., Predicting *in vivo* Mechanism of Action for Drug Leads using NMR Metabolomics, ACS Chemical Biology, **7**(1), 166-171 (2012). Reprinted with permission, copyright 2011 by American Chemical Society.

baumanii, *Pseudomonas aeruginosa*, and *Enterobacter* species that are infecting the majority of U.S. hospitals [6]. Also extreme drug resistant strains of *Mycobacterium tuberculosis* are a rising threat in the world.

The Infectious Diseases Society of America (IDSA) has proposed an initiative to develop and approve 10 novel antibiotics by the year 2020 [7]. However, existing drug discovery strategies may not be able to meet these challenges. Drug discovery programs rely heavily on target based high throughput screening (HTS) of large chemical libraries followed by lead optimization [8, 9]. Unfortunately, this approach has demonstrated an extremely high rate of failure and erroneous leads. Even when a valid HTS hit is found, it is uncertain if this chemical lead can penetrate into the bacterial cell while demonstrating *in vivo* activity.

NMR Metabolomics is evolving as a significant component of the drug discovery process and offers an inexpensive route to help overcome the multiple challenges faced by researchers [10]. Metabolomics is a relatively new field and is based on the identification and quantification of small molecules found in living cells or biofluids [11]. Since small molecules are downstream products of biomolecular process, the identity and concentration of metabolites provide biochemical signatures for tracking the physiological effects of antibiotic efficacy, selectivity, and toxicity. Characterizing these biochemical signatures relies upon the global determination of numerous endogenous small molecule followed by pattern recognition using multivariate analysis [12]. Such comprehensive biochemical information can be readily obtained using ^{1}H NMR spectroscopy with minimal sample handling while providing highly reproducible data in

an automated fashion [10]. Multivariate statistical analysis, such as orthogonal partial least-squares discriminant analysis (OPLS-DA), is typically employed to extract information from the large and complex NMR datasets [13]. Simply, OPLS-DA is used to identify clustering patterns from the major variations between NMR spectra [10].

Herein, we describe a new method using ^{1}H NMR and OPLS-DA to profile the *in vivo* mechanism of action of known antibiotics used to treat *M. tuberculosis*. More importantly, we aim to use the information to classify compounds with unknown mechanisms of action, but demonstrated anti-tubercular activity. Our approach is predicated on the hypothesis that drugs with similar modes of activity or therapeutic targets will have a similar impact on the metabolome of *M. smegmatis* and will cluster together in an OPLS-DA scores plot. Thus the mode of action of a novel chemical lead can be inferred from its clustering in an OPLS-DA scores plot relative to drugs with defined biological targets. Importantly, if the chemical lead is separated from known drugs in the OPLS-DA scores plot, then this result would infer a new mechanism of action and a potentially valuable, new antibiotic.

4.2 Methods

4.2.1 Determining Optimal Drug Dosage for NMR Metabolomics Experiments

M. smegmatis mc^{2}155 cells were grown at 37 °C with shaking at 200 rpm in 50 mL of Middlebrook 7H9 media until an average optical density at 600 nm ($O.D._{600}$) of 0.6 was achieved. Each drug was titrated over a concentration range of 1 to 24 times the literature MIC values and the cells were grown for an additional 2 hours. The optical

density was recorded and the growth rate inhibition was calculated by comparing the optical density of the treated cells to the untreated cells in the 2 hour time period. The desired drug dosage was determined where a drug concentration inhibits cell growth by approximately 50% of the growth rate of untreated cells.

4.2.2 Sample Preparation

A total of 190 *M. smegmatis* mc^2155 cultures were grown in 50 mL of Middlebrook 7H9 at 37 °C with shaking at 200 rpm until an $O.D._{600}$ of 0.6 was achieved. A total of 40 untreated cultures were used as a control and 10 cultures were inoculated with each antibiotics at the optimal dosage needed to inhibit cell growth by ~50% as described in Table 3.1. The cells were then grown for an additional 2 hours. The used media was removed and the cells were washed 3 times and re-suspended with 1 mL ice cold double distilled water. The cells were lysed using a FastPrep-24 instrument for 60 seconds at 6 m/s, and the supernatant was extracted and frozen in a dry ice ethanol bath. The samples were lyophilized and then re-suspended with 700 µL of 99.8% D_2O solution containing 50 mM phosphate buffer (pH 7.2, uncorrected) and 50 µM 3-(trimethylsilyl)propionic acid-2,2,3,3-d^4 (TMSP-d^4) as an internal standard for chemical shift referencing. The samples were then centrifuged for 5 minutes to remove any insoluble material, and 600 µL of the supernatant was transferred to an NMR tube.

Table 4.1. Description of Antimicrobial Compounds and Dosages Used in This Study

Compound	Class	Mechanism of Action	MIC[a] (ug/mL)	Dosage[b] (ug/mL)
Ampicillin	**P**enillins	Inhibits transpeptidation and prevents cell wall formation	16.0[c]	96.0
Cloramphenicol	Amphetamines	Inhibits protein synthesis by binding to the 50S ribosomal subunit	6.0	6.0
Ciprofloxacin	Fluoroquinolones	Inhibits DNA gyrase and prevents DNA supercoiling	0.2	2.0
D-cycloserine	Oxazolidinones	Inhibits alanine racemase and alanine ligase and prevents cell wall formation (different from other oxazolidinones that inhibit protein synthesis)	75.0	75.0
Ethambutol	Amino Alcohols	Disrupts arabinogalactan formation by inhibiting arabinosyl transferase	10.0	100.0
Ethionamide	Pyridine Derivatives	Inhibits mycolic acid formation similar to isoniazid	20.0	160.0
Gentamicin	Aminoglycosides	Inhibits protein synthesis by binding to the 30S ribosomal protein S12 and 16S rRNA	2.0	8.0
Isoniazid	Pyridine Derivatives	A prodrug that inhibits InhA and prevent mycolic acid synthesis	2.0	48.0
Kanamycin	Aminoglycosides	Inhibits protein synthesis by binding to the 30S ribosomal protein S12 and 16S rRNA	4.0	4.0
Rifampicin	Rifampicins	Inhibits RNA polymerase and prevent RNA synthesis	30.0	60.0
Streptomycin	Aminoglycosides	Inhibits protein synthesis by binding to the 30S ribosomal protein S12 and 16S rRNA	0.25	1.5
Vancomycin	Glycopeptides	Binds to the D-alanyl-D-alanine dipeptide and prevents cell wall formation	50	450
Amiodarone	Benzofurans	Unknown	26.6	212.8
Clofazimine	Anilines	Unkown	0.32	3.84
Chlorprothixene	Thioxanthines	Unknown	36.0	216.0

[a]Literature values of minimum inhibitory concentrations against *M. smegmatis* used as a starting point to determine an optimal dosage for the NMR metabolomics study. [b]Actual dosage used to treat *M. smegmatis* cells to inhibit growth by ~50% following drug treatment. [c]Reported for *M. smegmatis* β-lactamase and ribosomal protein S12 mutants.

4.2.3 NMR Data Collection and Processing

The NMR spectra were collected on a Bruker 500 MHz Avance spectrometer equipped with a triple resonance and z axis gradient cryoprobe. A BACS-120 sample changer was used for automated data collection. 1D ^{1}H NMR spectra were collected using excitation sculpting to remove the solvent signal and maintain a flat spectral baseline [14]. A total of 32K data points with a spectral width of 5482.5 Hz, 16 dummy scans and 128 scans were used to obtain each spectrum. The data was processed automatically using ACD/1D NMR Manager (Advanced Chemistry Development). Intelligent bucketing was used to integrate each spectral region with a bin size of 0.025 ppm. Each NMR spectrum was center averaged to minimize any experimental variations between bacterial cultures as follows [15],

$$Z = \frac{X_i - \bar{X}}{\sigma} \tag{4.1}$$

where $\bar{X}$ is the average signal intensity, σ is the standard deviation in the signal intensity, and X_i is the signal intensity within a bin. Noise regions of the spectra were omitted from the PCA analysis by setting the corresponding bins to zero [16]. OPLS-DA and PCA was performed using Simca-11.5+ (Umetrics), where each ^{1}H NMR spectra was reduced to a single point in the 2D OPLS-DA and PCA scores plot. The OPLSDA was calculated with two classes, untreated versus drug treated cell cultures, for the Y matrix with the NMR data incorporated into the X matrix. The OPLS-DA model was cross validated using a modified version of the leave-one-out technique, where 1 out of every 7 samples (spectra) were left out to calculate a model and predict the left out data [17]. The procedure was sequentially repeated leaving out a different 1/7th of the data. The predicted data was

then compared to the original data, where the quality assessment (Q^2) score provides a qualitative measure of the predictability of the model based on the consistency between the predicted and original data. An ideal value for Q^2 is one, where a typical value for a biological model is ≥ 0.4.

Metabolomic Tree diagrams with corresponding bootstrap values were created using our PCAtoTree program to interpret the OPLS-DA clustering pattern [18]. The metabolomics tree diagram is based on the Euclidean distances between the cluster centers from the 2D OPLS-DA scores plot. Standard bootstrapping techniques are used to generate a set of 100 distance matrices by randomly re-sampling the cluster centers and Euclidean distances. The set of 100 distance matrices are then used by PHYLIP (http://www.phylip.com) [19], phylogenetic software package, to generate 100 tree diagrams and a consensus tree diagram. The bootstrap numbers on the consensus tree diagram indicates the number of times each node was present in the set of 100 tree diagrams, where a bootstrap number below 50% indicates a generally insignificant node or insignificant separation between the clusters.

Four additional OPLS-DA models were generated to identify specific metabolites associated with drug activity: (i) inhibition of translation, transcription or DNA supercoiling drug treated cells versus untreated cells, (ii) inhibition of mycolic acid synthesis drug treated cells versus untreated cells, (iii) inhibition of cell wall synthesis drug treated cells versus untreated cells, and (iv) the three TAACF compounds versus untreated cells. S-plots and loading plots were generated from each OPLS-DA model. Bins (chemical shift values) demonstrating a covariance of greater than 0.10 or less than -0.10 were identified as major contributors to the class separation. Metabolites were

identified from this list of chemical shifts using the Human Metabolome Database (HMDB, http://www.hmdb.ca/) [20] with a chemical shift tolerance of 0.02 ppm. Metabolic network maps were then generated using Cytoscape (http://www.cytoscape.org/) [21] with the MetScape [22] plugin for the top 100 metabolite predicted by HMDB. Metabolites were excluded if not part of a network or not present in *M. smegmatis*.

4.3 Results and Discussion

Our methodology was demonstrated using 12 antibiotics known to inhibit the growth of *M. tuberculosis* and *M. smegmatis* (Table 4.1). The mechanism of action for each antibiotic was identified from the Drug Bank Database [23], and the minimum inhibitory concentrations (MIC) were obtained from the scientific literature [24-32]. In addition three chemical leads were randomly selected from the Tuberculosis Antimicrobial Acquisition and Coordinating Facility (TAACF) library of compounds (http://www.TAACF.org). The compounds were screened against *M. tuberculosis* and have comparable MICs to known TB drugs, but the biological target or mechanism of action was not reported by TAACF. The non-pathogenic *M. smegmatis* was used as a model system for the NMR metabolomics study.

In order to analyze changes in the *M. smegmatis* metabolome, the drug dosage needs to be below lethal levels and only affect cell growth. Typically, a drug concentration that inhibits cell growth by approximately 50% of the growth rate of untreated cells is desirable. While MIC values are available from the literature, these concentrations are based on standardized drug gradients, inoculum sizes, and readout

endpoints. Additionally, the reported MIC's were obtained with different bacterial strains, at different growth stages or cell densities, and under a variety of experimental conditions that includes either broth or agar methods. Further complicating the situation is the diversity of MIC's values reported for a single drug. Thus, the literature MIC values listed in Table 4.1 were simply used as a starting point to determine an optimal dosage for the NMR metabolomics study under our experimental conditions. Each drug was titrated over a concentration range of 1 to 24 times the literature MIC values. The individual drug concentrations needed to achieve ~50% growth inhibition are reported in Table 4.1. An average growth inhibition of 43.1±10.5% was observed after the addition of each of the 15 drugs. Preparation of the *M. smegmatis* cell cultures for metabolomics analysis was then performed using the optimal dosage for each drug.

Due to the inherit variability of biological samples and to provide a robust statistical analysis, 10 cultures inoculated with each antibiotic and 40 cultures of untreated cells were prepared for the NMR metabolomics study. A 1D ^{1}H NMR spectrum was collected for each biological sample, which were normalized using center averaging and analyzed using OPLS-DA. A representative 2D OPLS-DA scores plot displaying a comparison between 6 antibiotics with known mechanisms of action is shown in Figure 4.1. The OPLS-DA model was cross-validated using a modified leave-one-out method. A quality assessment score (Q^2) of 0.82 was obtained, which is an excellent result compared to an ideal score of 1.00. The, the cross-validation indicates a highly reliable model. Each point in the 2D OPLS-DA scores plot represents a single 1D ^{1}H NMR spectrum of a specific drug treated or untreated cell culture. The 2D OPLS-DA

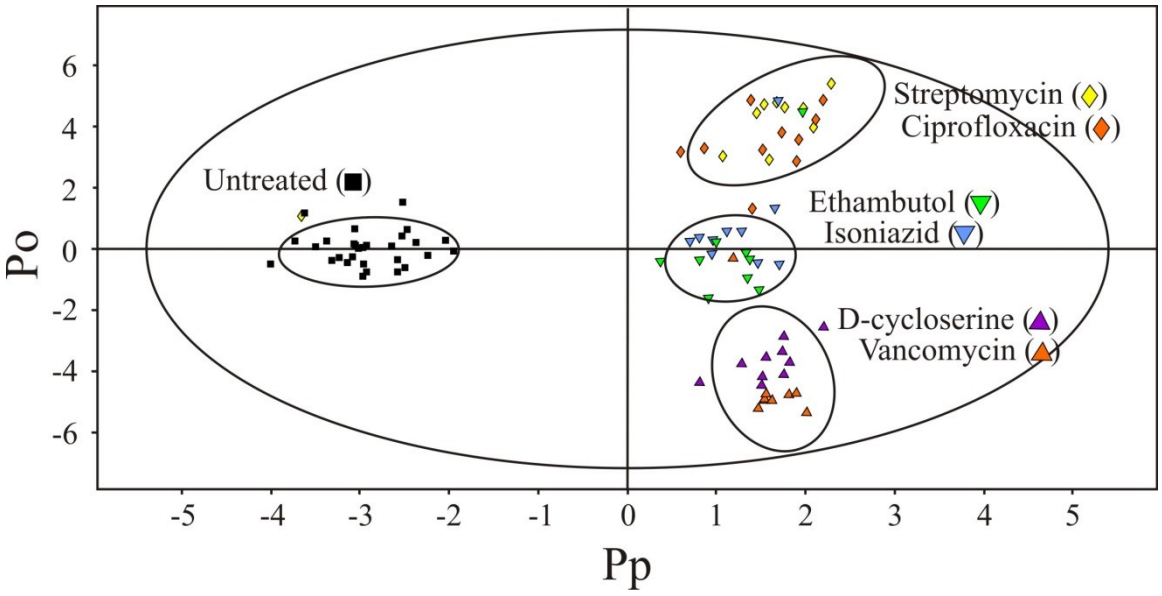


Figure 4.1. 2D OPLS-DA scores plot demonstrating the clustering pattern obtained for six different antibiotics with known and distinct biological targets: untreated *M. smegmatis* cells, ciprofloxacin, streptomycin, ethambutol, isoniazid, D-cycloserine, and vancomycin treated *M. smegmatis* cells. Symbols and labels are indicated on the plot. The ellipses correspond to the 95% confidence limits from a normal distribution for each cluster. The untreated *M. smegmatis* cells were designated the control class, and the remainder of the cells was designated as treated. The OPLS-DA used one predictive component and three orthogonal components to yield a R^2X of 0.610, R^2Y of 0.893, and Q^2 of 0.82.

scores plot consists of 4 separate clustering patterns, which demonstrates that each group has a considerably different impact on the metabolome of *M. smegmatis*.

Importantly, all the drug-treated *M. smegmatis* cell cultures form distinct and separate clusters from the untreated cell cultures. This is consistent with all the drugs being biologically active and inhibiting *M. smegmatis* cell growth. Antibiotics that share a similar or identical biological target were observed to cluster together in the OPLS-DA scores plot. For example, ethambutol and isoniazid inhibit mycolic acid biosynthesis that prevents the formation of the arabinogalactan-mycolic acid matrix. Streptomycin and ciprofloxacin form the second cluster. Streptomycin prevents protein synthesis and ciprofloxacin inhibits DNA supercoiling that affects replication, transcription, and repair, leading to a similar disruption in protein synthesis. Since these two antibiotics cluster together, it implies that the inhibition of transcription or translation results in a similar impact on the metabolome. Vancomycin and D-cycloserine both affect cell wall formation and form the third cluster. In a principal component analysis (PCA) of the data (Figure 4.2) there is a more pronounced separation between vancomycin and D-cycloserine along PC2. This reflects a fundamental difference between PCA and OPLS-DA, where PCA is limited to a linear model and does not readily differentiate between within-class and between-class variations [13, 33]. Correspondingly, OPLS-DA is preferred as long as cross-validation verifies a reliable model.

The NMR metabolomics analysis was then expanded to include a total of 12 drugs with known biological targets and 3 compounds randomly chosen from the TAACF library. Amiodorone, clofazamine and chlorprothixene are active against TB, but have unknown mechanisms of action according to the TAACF database. Nevertheless, the

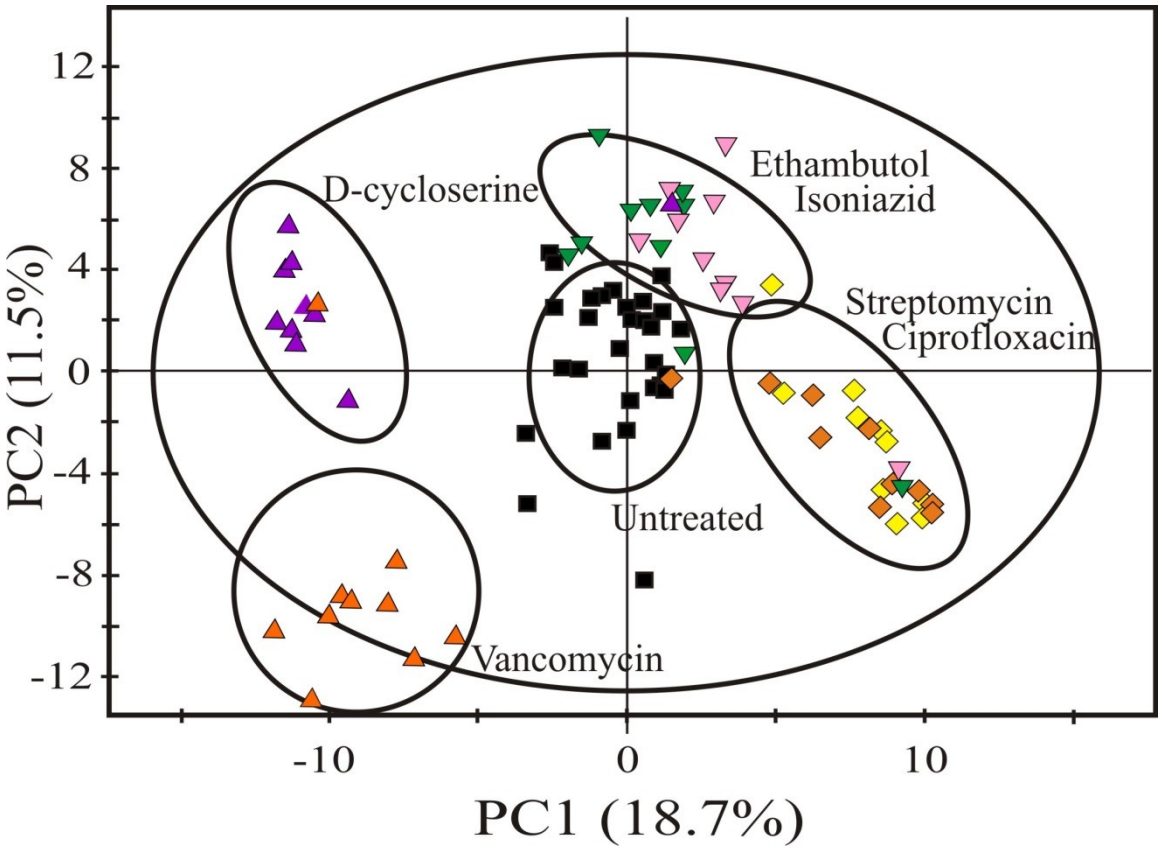


Figure 4.2. PCA scores plot demonstrating the clustering pattern of 6 different classes of known antibiotics: untreated *M. smegmatis* cells (■), ciprofloxacin (♦), streptomycin (♦), ethambutol (▼), isoniazid (▼), D-cycloserine (▲), and vancomycin (▲) treated *M. smegmatis* cells. The ellipses correspond to the 95% confidence limits from a normal distribution for each cluster.

three compounds are known drugs, where amiodorone is an antiarrhythmic agent that affects potassium efflux, chlorprothixene is an antipsychotic drug that inhibits dopamine receptors, and clofazamine is a 40 year-old leprosy treatment with an ill-defined biological activity. The 2D OPLS-DA scores plot (Figure 4.3a) identified 4 distinct clusters and yielded a highly reliable cross validation Q^2 score of 0.671. As before, the different clusters are correlated with distinct modes of action: inhibition of cell wall formation, inhibition of mycolic acid biosynthesis, and inhibition of transcription, translation or the overall effects of DNA supercoiling.

The accompanying metabolomics tree diagram (Figure 4.3b) clearly visualizes the relative groupings of the three antibiotic classes [18]. The bootstrap numbers of 89 to 100 indicate a statistically significant separation between the five clusters and the reliability of the general drug and TAACF classifications. The metabolic tree diagram also provides a finer separation between the drugs within each cluster. These separations may reflect actual differences in the specific drug targets. For example, D-cycloserine and vancomycin are on separate branches in the cell wall node potentially because D-cycloserine inhibits alanine racemase and alanine ligase compared to vancomycin binding the D-alanyl-D-alanine dipeptide. Alternatively, the separation may result from differences in the relative activity of the drug. Streptomycin forms a separate branch in the transcription, translation or DNA supercoiling drug cluster despite having a similar target (binding to the 30S ribosomal protein S12 and 16S rRNA) relative to other members within the cluster. But, streptomycin is one of the most active compounds

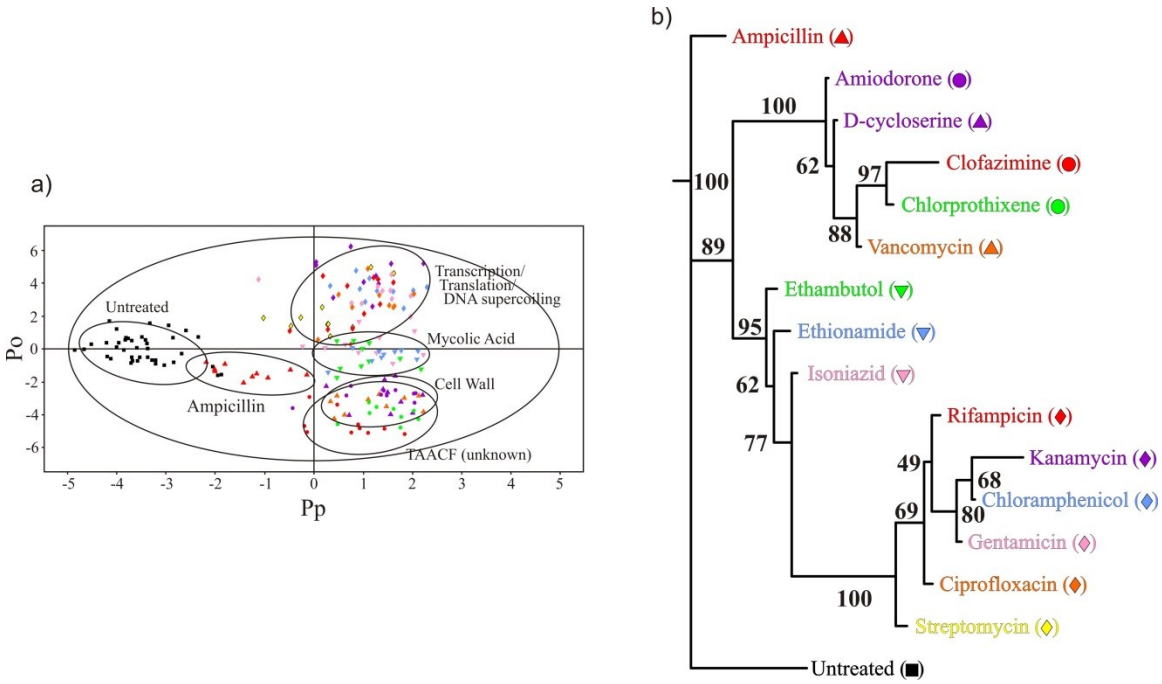


Figure 4.3. a) 2D OPLS-DA scores plot demonstrating the clustering pattern for 12 antibiotics with known biological targets and three compounds of unknown *in vivo* activity: untreated *M. smegmatis* cells, chloramphenicol, ciprofloxacin, gentamicin, kanamycin, rifampicin, streptomycin, ethambutol, ethionamide, isoniazid, ampicillin, D-cycloserine, vancomycin, amiodorone, chlorprothixene, and clofazimine treated *M. smegmatis* cells. The symbols correspond with the coloring scheme and labeled symbols indicated on the tree diagram in (b). The ellipses correspond to the 95% confidence limits from a normal distribution for each cluster. The untreated *M. smegmatis* cells (black square) was designated the control class, and the remainder of the cells were designated as treated. The OPLS-DA used one predictive component and six orthogonal components to yield a R^2X of 0.715, R^2Y of 0.803, and Q^2 of 0.671. **b)** Metabolomics tree diagram determined from the OPLS-DA scores plot. The coloring scheme and associated symbol for each compound in the tree diagram correlates with colored symbols in the OPLS-DA scores plot. The bootstrap numbers for each node are indicated on the tree diagram.

tested, requiring only a dosage of 1.5 μg/ml to inhibit *M. smegmatis* growth by approximately 50%. Also, over-interpreting these subtle separations may be erroneous since the within cluster differences may simply reflected experimental variability and may not be biologically relevant. For instance, an average growth inhibition of 43.1 ± 10.5% was observed after the addition of each of the 15 drugs. This dosage variability may lead to unintended separations in the 2D OPLS-DA scores plot. Essentially, the reliability of these finer cluster differences is dependent on additional supportive biological data.

Surprisingly, amiodorone, chlorprothixene, and clofazamine were found to cluster together in the 2D OPLS-DA scores plot and metabolic tree diagram. This was an unexpected result given that the three compounds were randomly selected from the large TAACF library and have diverse therapeutic usages. But, it also implies the three compounds share a similar mechanism of action in TB. Importantly, the three TAACF compounds also cluster with the antibiotics that disrupt cell wall formation, ampicillin, D-cycloserine and vancomycin. This infers a similar mode of action between the three TAACF compounds and the antibiotics that are known to interfere with bacterial cell walls. A subsequent literature search indicated that the three TAACF compounds have been previously shown to disrupt bacterial membranes in organisms distinct from TB [34-38]. Thus, the literature results are consistent with our NMR metabolomics analysis, which support our general classification of the TAACF compounds as interfering with bacterial cell walls. It is important to note that while ampicillin is a member of this class of antibiotics, it is also skewed toward the untreated cells in the 2D OPLS-DA scores plot. Presumably, this is because of *M. smegmatis* β-lactamase activity that provides

resistance to ampicillin [32, 39]. The impact of ampicillin on the metabolome of *M. smegmatis* is significantly diminished such that ampicillin *M. smegmatis* is similar to untreated cells. As described previously, there are some differences between the OPLS-DA and PCA scores plot (Figure 4.4). There is less discrimination between the untreated and drug treated cells in the 2D PCA scores plot. This is not unexpected since PLS is preferred over PCA for discrimination been classes [33]. Also, there is a separation between the three TAACF compounds and the cell wall antibiotics in the PCA scores plot, but the TAACF compounds are still closer to the cell wall antibiotics in the associated metabolic tree diagram (Figure 4.4). In fact, the OPLS-DA and PCA metabolomic tree diagrams are quite similar despite these visible differences in the scores plots. Additionally, the quality of the OPLS-DA model is apparent from the fit to the data, $R^2X > 0.610$ and $R^2Y > 0.803$, and the reliability of the model is apparent from the cross-validation Q^2 score > 0.617. Further validation of the OPLS-DA drug and TAACF classifications comes from the analysis of the metabolites identified as the major contributors to the OPLS-DA class separation (Figures 4.5-4.8 and Tables 4.2-4.5).

The S-plots and loading plots determined from the OPLS-DA models identify the chemical shifts (and associated metabolites) that contribute to the observed separation between the untreated and treated cells in the 2D OPLS-DA scores plot. The metabolites and corresponding pathways predominately perturbed by the addition of each drug class are listed in supplemental Tables 4.2-4.5. While there are some broad similarities in the metabolites and pathways affected by the drugs because the comparisons are all made relative to untreated cells, there are also some distinct differences. For example, proline, cytidine, uridine and inosine (pyrimidine and purine pathways) are all uniquely decreased

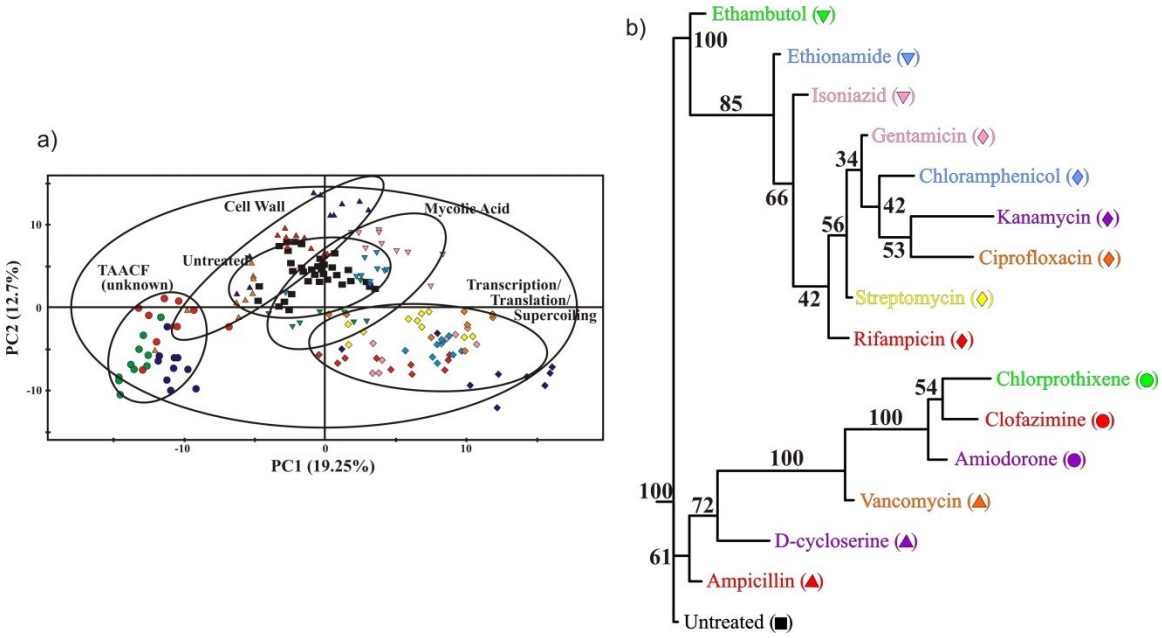


Figure 4.4. a) PCA scores plot demonstrating clustering patterns of 12 antibiotics with known mechanisms and 3 compounds that are unknown: untreated *M. smegmatis* cells, chloramphenicol, ciprofloxacin, gentamicin, kanamycin, rifampicin, streptomycin, ethambutol, ethionamide, isoniazid, ampicillin, D-cycloserine, vancomycin, amiodorone, chlorprothixene, and clofazimine treated *M. smegmatis* cells. The symbols correspond with the coloring scheme and labeled symbols indicated on the tree diagram in (b). The ellipses correspond to the 95% confidence limits from a normal distribution for each cluster. **b)** Tree diagram of the PCA scores plot. The coloring scheme for each compound in the tree diagram represents the data points in the scores plot. The bootstrap numbers for each node are indicated on the tree diagram.

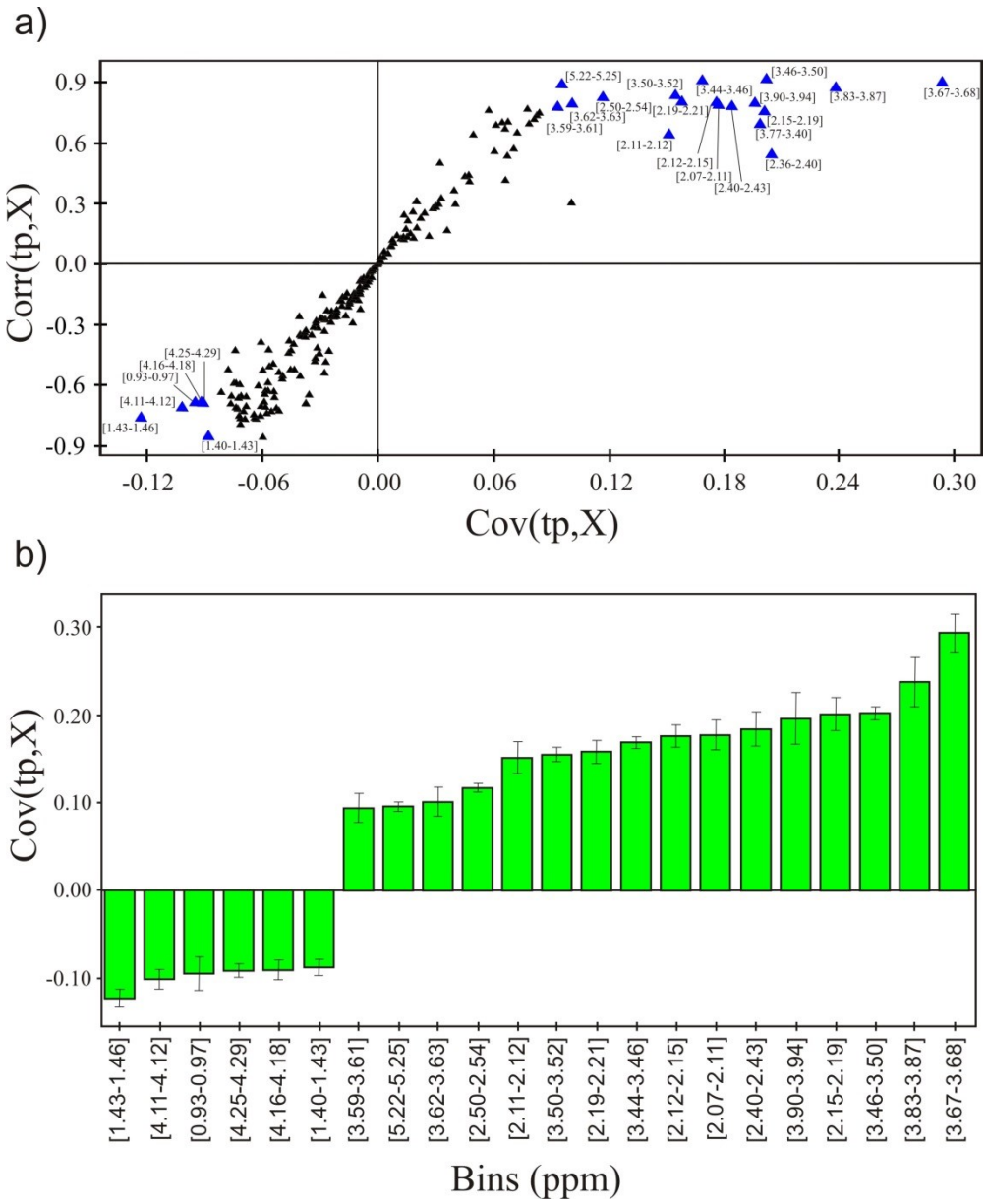


Figure 4.5: a) OPLS-DA S-plot comparing the *M. smegmatis* treated with antibiotics known to inhibit transcription or translation and untreated cell cultures. Each point in the S-plot represents a specific bin containing a chemical shift range of about 0.25ppm, and the range varied by 50% using intelligent binning in the ACD 1D NMR processor. **b)** OPLS-DA loading plot comparing the *M. smegmatis* treated with antibiotics known to inhibit transcription or translation and untreated cell cultures.

Table 4.2. Metabolites perturbed from drugs affecting transcription, translation or DNA supercoiling.

Pathway	Metabolite	HMDB Score[a]	Relative Change[b]
Glycolysis	Glucose-6-phosphate	19/50	+
	Fructose-6-phosphate	15/37	+
	Glycerol-3-phosphate	18/34	+
	Lactate	2/2	-
Aminosugar	N-acetyl-D-glucosamine	23/50	+
	N-acetyl-Neuraminate	12/37	+
Folate metabolism	Folate	10/33	+
Glycine,Serine,Threonine metabolism	Serine	5/12	+
Cysteine, Methionine metabolism	Methionine	18/22	+
TCA	Citrate	2/4	+
	Isocitrate	6/16	+
	Acetoacetate	1/2	+
Branched Chain Amino Acids	Isoleucine	7/42	-
	Leucine	5/25	-
Alanine, Aspartate, Glutamate metabolism	Alanine	2/6	-
	Glutamate	12/30	+
	Glutamine	16/28	+
Lysine Biosynthesis	Lysine	5/41	-
Proline Metabolism	Proline	8/44	-
Pyrimidine	Cytidine	7/25	-
	Uridine	5/26	-
Purine	Inosine	4/21	-

[a]Number of peaks in query that matches the number of peaks in the Human Metabolome Database. Peaks in the query can be matched multiple times. [b]Relative change of metabolite concentration in drug treated cultures compared to untreated cultures.

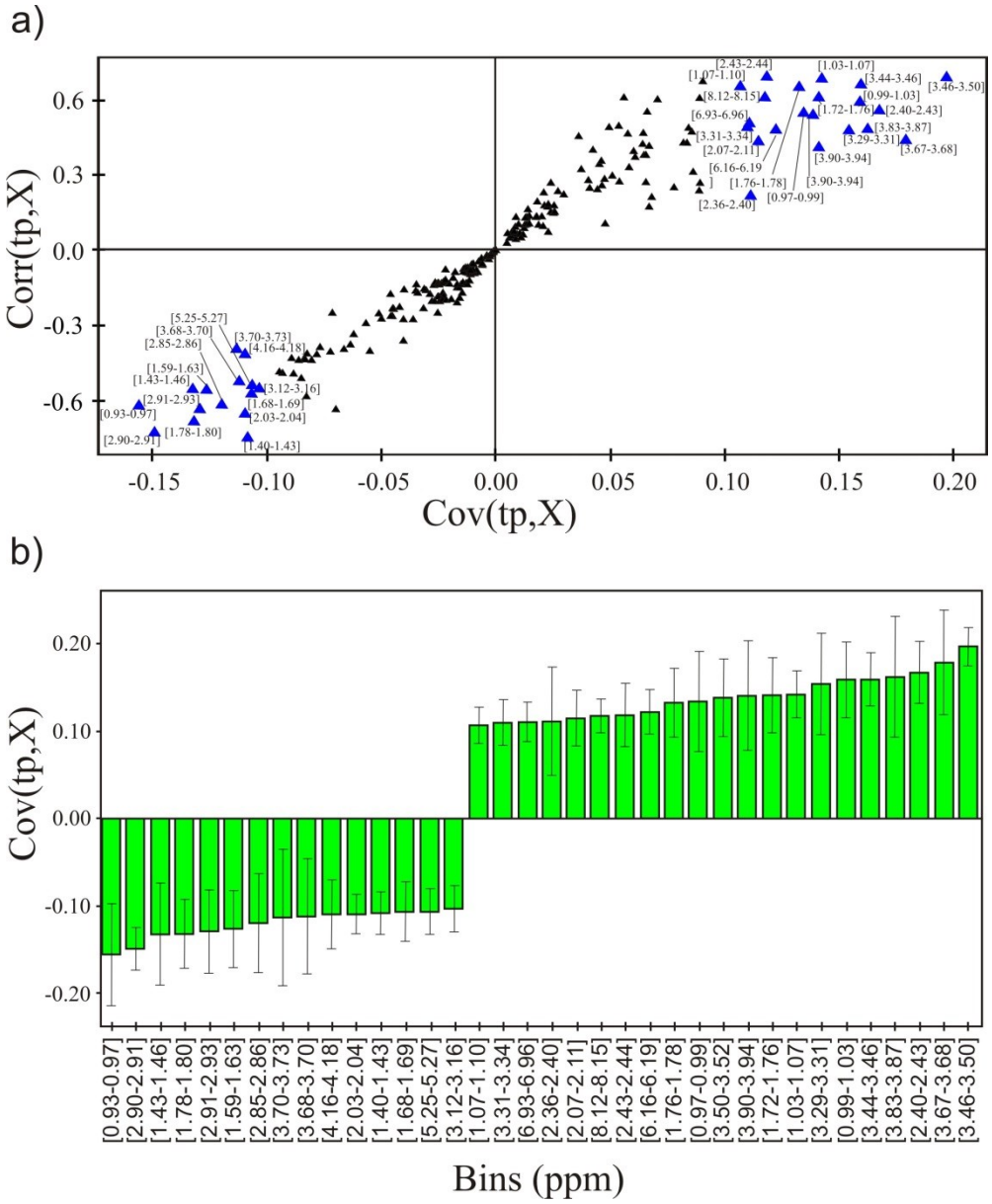


Figure 4.6. a) OPLS-DA S-plot comparing the *M. smegmatis* treated with antibiotics known to inhibit mycolic acid pathways and untreated cell cultures. **b)** OPLS-DA loading plot comparing the *M.smegmatis* treated with antibiotics known to inhibit mycolic acid pathways and untreated cell cultures.

Table 4.3. Metabolites perturbed from drugs affecting mycolic acid pathways

Pathway	Metabolite	HMDB Score[a]	Relative Change[b]
Glycolysis	Glucose-6-phosphate	16/50	+
	Fructose-6-phosphate	15/37	+
	Oxaloacetate	1/1	+
	Lactate	2/6	+
TCA	Acetoacetate	1/2	+
	2-Oxoglutarate	3/6	+
	Isocitrate	6/16	+
Aminosugar	Arabinose	15/40	+
Alanine, Aspartate, Glutamate metabolism	GABA	6/11	+
	Glutamate	12/30	+
	Asparagine	5/12	-
	Alanine	2/6	-
Glycine, Serine, Threonine metabolism	2-Ketobutyrate	3/7	+
	Homoserine	12/44	-
Branched Chain metabolism	Valine	2/16	+
	Leucine	3/25	+
Pyrimidine	Orotate	1/1	+
	Ureidopropionate	6/7	+
Glycerophospholipid	Choline phosphate	8/15	-
Lysine Biosynthesis	Lysine	15/47	-
Arginine and Proline metabolism	Spermidine	7/21	-
	Citrulline	16/44	-
Ascorbate metabolism	Ascorbate	6/16	-
	Dehydroascorbate	10/35	-

[a]Number of peaks in query that matches the number of peaks in the Human Metabolome Database. Peaks in the query can be matched multiple times. [b]Relative change of metabolite concentration in drug treated cultures compared to untreated cultures.

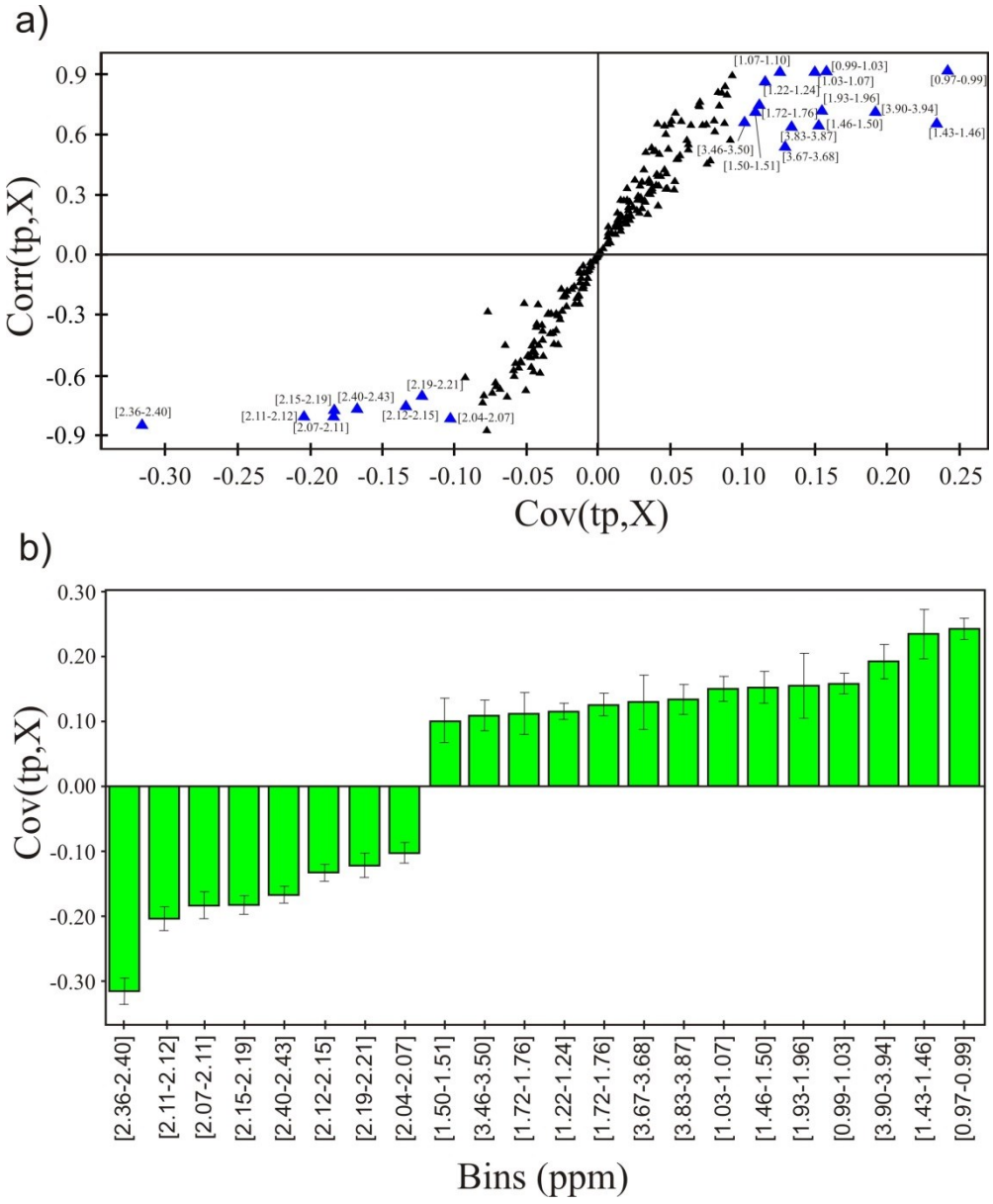


Figure 4.7. a) OPLS-DA S-plot comparing the *M. smegmatis* treated with antibiotics known to inhibit cell wall synthesis and untreated cell cultures. **b)** OPLS-DA loading plot comparing the *M. smegmatis* treated with antibiotics known to inhibit cell wall formation and untreated cell cultures.

Table 4.4. Metabolites perturbed from drugs affecting cell wall.

Pathway	Metabolite	HMDB Score[a]	Relative Change[b]
Glycolysis	Glucose-6-phosphate	19/50	+
	Fructose-6-phosphate	20/37	+
	Oxaloacetate	1/1	+
Aminosugar	N-acetyl-D-glucosamine	25/50	+
	N-acetylneuraminate	12/37	+
Alanine, Aspartate, Glutamate metabolism	Alanine	2/6	+
	GABA	8/11	+
	Glutamine	18/28	+
	Glutamate	18/30	+
Branched Chain Amino Acids	Valine	10/16	+
	Isoleucine	31/42	-
Glycine, Serine, Threonine metabolism	Serine	7/12	-
	2-ketobutyrate	3/7	+
Cysteine, Methionine metabolism	Methionine	15/22	-
Lysine Biosynthesis	Lysine	15/47	+
Folate metabolism	Folate	13/33	+

[a]Number of peaks in query that matches the number of peaks in the Human Metabolome Database. Peaks in the query can be matched multiple times. [b]Relative change of metabolite concentration in drug treated cultures compared to untreated cultures.

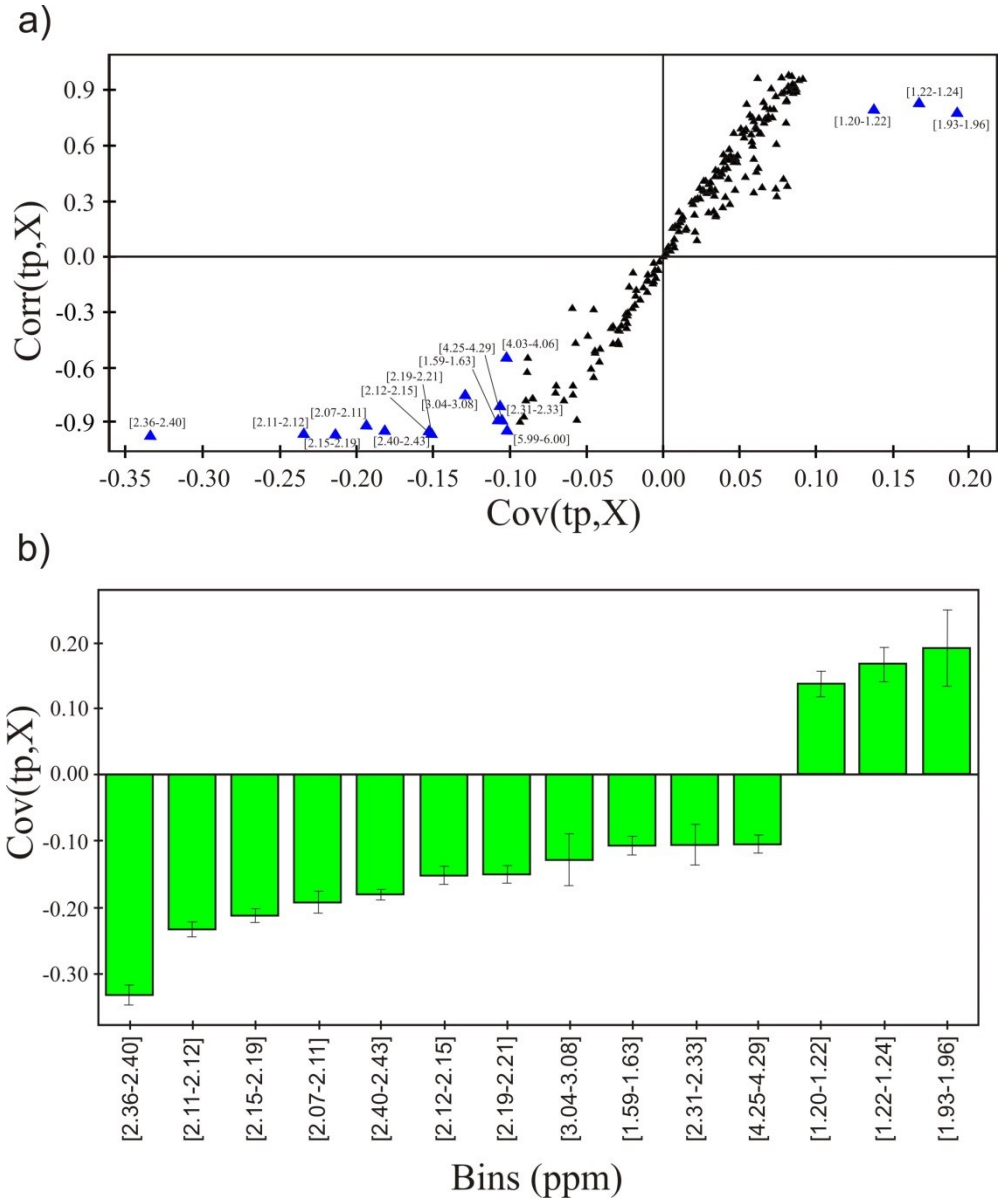


Figure 4.8: a) OPLS-DA S-plot comparing the *M. smegmatis* treated with antibiotics with unknown mechanism and untreated cell cultures. **b)** OPLS-DA loading plot comparing the *M. smegmatis* treated with antibiotics with unknown mechanism and untreated cell cultures.

Table 4.5. Metabolites perturbed from drugs of unknown mechanism.

Pathway	Metabolite	HMDB Score[a]	Relative Change[b]
Glycolysis	Oxaloacetate	1/1	-
Alanine, Aspartate, Glutamate metabolism	Glutamine	20/28	-
	Glutamate	25/30	-
Branched Chain Amino Acids	Isoleucine	20/42	+
Cysteine, Methionine metabolism	Methionine	15/22	-
Folate metabolism	Folate	24/33	-

[a]Number of peaks in query that matches the number of peaks in the Human Metabolome Database. Peaks in the query can be matched multiple times. [b]Relative change of metabolite concentration in drug treated cultures compared to untreated cultures.

by drugs that affect transcription, translation or DNA supercoiling. Obviously, nucleotides are essential metabolic precursors to DNA and RNA synthesis. Alternatively, choline phosphate, lysine, spermidine, citrulline, ascorbate and dehydroascorbate (glycerophospholipid, lysine biosynthesis, arginine and proline metabolism, and ascorbate metabolism pathways) are decreased by drugs affecting the mycolic acid pathway. Ascorbate metabolism is directly linked to the mycolic acid pathway, where ascorbate leads to arabinose. Arabinose is the primary precursor for the arabinogalactan-mycolic acid complex. Also, the inhibition of spermidine synthesis has been previously observed for drugs targeting the mycolic acid pathway in mycobacteria [40]. Importantly, the set of metabolites affected by the TAACF compounds were identical to metabolites perturbed by D-cycloserine and vancomycin. Both show a decrease in oxaloacetate, glutamine, glutamate, methionine and folate and an increase in isoleucine. Clearly, amino acids and their precursors are important components in peptidoglycan, cell wall and cell membrane synthesis. There were some additional metabolites that are increased by the addition of D-cycloserine and vancomycin that were not observed with the TAACF compounds. These include other amino acids (alanine, lysine serine, valine) and other precursors to peptidoglycan synthesis (N-acetyl-D-glucosamine, N-acetylneuraminate). Overall, the identity of the specific metabolites perturbed by each drug class is consistent with the 2D OPLS-DA scores plot clustering pattern and drug classifications.

4.4 Conclusion

In conclusion, we have demonstrated that different classes of antibiotics uniquely affect the metabolome of *M. smegmatis.* These metabolomic changes are directly correlated with broad mechanisms of action that are associated with each TB class of antibiotics, disruption of cell walls or membranes, inhibition of transcription, translation or DNA supercoiling, or the inhibition of mycolic acid biosynthesis. Thus, NMR metabolomics provides an efficient, simple and unbiased approach for providing rapid classification of promising drugs leads that emerge from HTS. This is critical since HTS does not provide any information on mechanisms of action; only relative activity with a high-false positive rate. Instead, the *in vivo* biological activity of a novel lead can be inferred by its relative clustering to existing drug classes in an OPLS-DA scores plot derived from metabolomics data. Importantly, a chemical lead that forms a distinct cluster from known drugs infers a potential new mechanism of action and a reason to prioritize the chemical lead for a detailed follow-up investigation. Of course, the induced metabolomic changes relative to untreated cells provide further confirmation of *in vivo* efficacy, which was implied from the HTS results. While the technique was demonstrated with *M. smegmatis*, it is generally applicable to bacterial pathogens and the effect of therapeutic agents on human cell lines in addition to the analysis of biofluids.

4.5 References

1. Jones, K.E., et al., *Global Trends in Emerging Infectious Diseases.* Nature, 2008. **451**(7181): p. 990-3.

2. Meya, D.B. and K.P. McAdam, *The TB Pandemic: An Old Problem Seeking New Solutions.* Journal of Internal Medicine, 2007. **261**(4): p. 309-29.

3. Gwynn, M.N., et al., *Challenges of Antibacterial Discovery Revisited.* Annals of the New York Academy of Sciences, 2010. **1213**: p. 5-19.

4. Talbot, G.H., et al., *Bad Bugs Need Drugs: An Update on the Development Pipeline from the Antimicrobial Availability Task Force of the Infectious Diseases Society of America.* Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America, 2006. **42**(5): p. 657-68.

5. Coates, A.R., G. Halls, and Y. Hu, *Novel Classes of Antibiotics or More of the Same?* British Journal of Pharmacology, 2011. **163**(1): p. 184-94.

6. Fischbach, M.A. and C.T. Walsh, *Antibiotics for Emerging Pathogens.* Science, 2009. **325**(5944): p. 1089-93.

7. *The 10 x '20 Initiative: Pursuing a Global Commitment to Develop 10 New Antibacterial Drugs by 2020.* Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America, 2010. **50**(8): p. 1081-3.

8. Brown, D., *Unfinished Business: Target-Based Drug Discovery.* Drug Discovery Today, 2007. **12**(23-24): p. 1007-12.

9. Bunnage, M.E., *Getting Pharmaceutical R&D Back on Target.* Nature Chemical Biology, 2011. **7**(6): p. 335-9.

10. Powers, R., *NMR Metabolomics and Drug Discovery.* Magnetic Resonance in Chemistry : MRC, 2009. **47 Suppl 1**: p. S2-11.

11. Baker, M., *Metabolomics: From Small Molecules to Big Ideas.* Nature Methods, 2011. **8**(2): p. 117-121.

12. Lindon, J.C., E. Holmes, and J.K. Nicholson, *Pattern Recognition Methods and Applications in Biomedical Magnetic Resonance.* Progress in Nuclear Magnetic Resonance Spectroscopy, 2001. **39**(1): p. 1-40.

13. Bylesjo, M., et al., *OPLS Discriminant Analysis: Combining the Strengths of PLS-DA and SIMCA Classification.* Journal of Chemometrics, 2006. **20**(8-10): p. 341-351.

14. Hwang, T.L. and A.J. Shaka, *Water Suppression That Works - Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients.* Journal of Magnetic Resonance Series A, 1995. **112**(2): p. 275-279.

15. Craig, A., et al., *Scaling and Normalization Effects in NMR Spectroscopic Metabonomic Data Sets.* Analytical Chemistry, 2006. **78**(7): p. 2262-7.

16. Halouska, S. and R. Powers, *Negative Impact of Noise on the Principal Component Analysis of NMR Data.* Journal of Magnetic Resonance, 2006. **178**(1): p. 88-95.

17. Westerhuis, J.A., et al., *Assessment of PLSDA Cross Validation.* Metabolomics, 2008. **4**(1): p. 81-89.

18. Werth, M.T., et al., *Analysis of Metabolomic PCA Data Using Tree Diagrams.* Analytical Biochemistry, 2010. **399**(1): p. 58-63.

19. Retief, J.D., *Phylogenetic Analysis Using PHYLIP.* Methods in Molecular Biology, 2000. **132**: p. 243-58.

20. Wishart, D.S., et al., *HMDB: A Knowledgebase for the Human Metabolome.* Nucleic Acids Research, 2009. **37**: p. D603-10.

21. Smoot, M.E., et al., *Cytoscape 2.8: New Features for Data Integration and Network Visualization.* Bioinformatics, 2011. **27**(3): p. 431-2.

22. Gao, J., et al., *Metscape: A Cytoscape Plug-in for Visualizing and Interpreting Metabolomic Data in the Context of Human Metabolic Networks.* Bioinformatics, 2010. **26**(7): p. 971-3.

23. Knox, C., et al., *DrugBank 3.0: A Comprehensive Resource for 'Omics' Research on Drugs.* Nucleic Acids Research, 2011. **39**: p. D1035-41.

24. Chung, G.A., et al., *High-Throughput Screen for Detecting Antimycobacterial Agents.* Antimicrobial Agents and Chemotherapy, 1995. **39**(10): p. 2235-8.

25. Stephan, J., et al., *Multidrug Resistance of a Porin Deletion Mutant of Mycobacterium smegmatis.* Antimicrobial Agents and Chemotherapy, 2004. **48**(11): p. 4163-70.

26. Malik, M., et al., *Lethality of Quinolones Against Mycobacterium smegmatis in the Presence or Absence of Chloramphenicol.* Antimicrobial Agents and Chemotherapy, 2005. **49**(5): p. 2008-14.

27. Lu, T., et al., *Effect of Chloramphenicol, Erythromycin, Moxifloxacin, Penicillin and Tetracycline Concentration on the Recovery of Resistant Mutants of Mycobacterium smegmatis and Staphylococcus aureus.* The Journal of Antimicrobial Chemotherapy, 2003. **52**(1): p. 61-4.

28. Halouska, S., et al., *Use of NMR Metabolomics to Analyze the Targets of D-Cycloserine in Mycobacteria: Role of D-Alanine Racemase.* Journal of Proteome Research, 2007. **6**(12): p. 4608-14.

29. Burguiere, A., et al., *Altered Expression Profile of Mycobacterial Surface Glycopeptidolipids Following Treatment with the Antifungal Azole Inhibitors Econazole and Clotrimazole.* Microbiology, 2005. **151**(Pt 6): p. 2087-95.

30. Mick, V., et al., *Transcriptional Analysis of and Resistance Level Conferred by the Aminoglycoside Acetyltransferase Gene aac(2')-Id from Mycobacterium smegmatis.* The Journal of Antimicrobial Chemotherapy, 2008. **61**(1): p. 39-45.

31. Chacon, O., et al., *Mycobacterium smegmatis D-Alanine Racemase Mutants Are Not Dependent on D-Alanine for Growth.* Antimicrobial Agents and Chemotherapy, 2002. **46**(1): p. 47-54.

32. Flores, A.R., L.M. Parsons, and M.S. Pavelka, Jr., *Characterization of Novel Mycobacterium tuberculosis and Mycobacterium smegmatis Mutants Hypersusceptible to Beta-Lactam Antibiotics.* Journal of Bacteriology, 2005. **187**(6): p. 1892-900.

33. Barker, M. and W. Rayens, *Partial Least Squares for Discrimination.* Journal of Chemometrics, 2003. **17**(3): p. 166-173.

34. Rosa, S.M., et al., *Amiodarone Interactions with Membrane Lipids and with Growth of Bacillus stearothermophilus Used as a Model.* Applied Biochemistry and Biotechnology, 2000. **87**(3): p. 165-75.

35. Oliva, B., et al., *Anti-Staphylococcal Activity and Mode of Action of Clofazimine.* The Journal of Antimicrobial Chemotherapy, 2004. **53**(3): p. 435-40.

36. Kristiansen, J.E., et al., *Non-Antibiotics Reverse Resistance of Bacteria to Antibiotics.* In vivo, 2010. **24**(5): p. 751-4.

37. O'Neill, A.J., et al., *Comparison of Assays for Detection of Agents Causing Membrane Damage in Staphylococcus aureus.* The Journal of Antimicrobial Chemotherapy, 2004. **54**(6): p. 1127-9.

38. Hurdle, J.G., et al., *Targeting Bacterial Membrane Function: an Underexploited Mechanism for Treating Persistent Infections.* Nature Reviews. Microbiology, 2011. **9**(1): p. 62-75.

39. Kwon, H.H., H. Tomioka, and H. Saito, *Distribution and Characterization of Beta-Lactamases of Mycobacteria and Related Organisms.* Tubercle and Lung Disease : The Official Journal of the International Union Against Tuberculosis and Lung Disease, 1995. **76**(2): p. 141-8.

40. Paulin, L.G., E.E. Brander, and H.J. Poso, *Specific Inhibition of Spermidine Synthesis in Mycobacteria spp. by the Dextro Isomer of Ethambutol.* Antimicrobial Agents and Chemotherapy, 1985. **28**(1): p. 157-9.

CHAPTER 5

TRICARBOXYLIC ACID CYCLE-DEPENDENT REGULATION OF STAPHYLOCOCCUS EPIDERMIS POLYSACCHARIDE INTERCELLULAR ADHESIN SYNTHESIS[§]

5.1 Introduction

Staphylococcus epidermidis is the most frequently isolated nosocomial pathogen from the blood cultures of adult and pediatric intensive care unit patients in the United States of America [1]. Frequently, *S. epidermidis* infections involve the formation of a biofilm on the surfaces of biomaterials (e.g., catheters) [2]. A biofilm is a complex aggregation of bacteria usually encapsulated by an adhesive exopolysaccharide matrix. The exopolysaccharide matrix provides structural stability to the biofilm, enhanced adhesion to surfaces, and protection from host defenses and antibiotics [3-5]. In *S. epidermidis*, the exopolysaccharide facilitating bacterial adherence in a biofilm is polysaccharide intercellular adhesin (PIA). PIA is an *N*-acetylglucosamine polymer [6] of at least 130 residues [7] whose synthesis requires the enzymes encoded by genes in the intercellular adhesin (*ica*) operon (*icaADBC*) [8]. The synthesis of PIA is required for the maturation of biofilms [8], important for evasion of the host innate immune system [5, 9], and critical in biomaterial-associated infections [10, 11].

[§] Chapter 5 was adapted from Sadykov, M., et. al., Tricarboxylic Acid Cycle-Dependent Regulation of *Staphylococcus epidermidis* Polysaccharide Intercellular Adhesin Synthesis, *Journal of Bacteriology* (2008), **190** (23), 7621-7642. Reprinted with permission, copyright 2008 American Society for Microbiology.

Bacteria live in environments subject to rapid changes in the availability of the nutrients necessary to provide energy and biosynthetic intermediates for the synthesis of macromolecules. Consequently, bacterial survival depends on the ability to regulate expression of genes coding for enzymes required for growth in the altered environment. In pathogenic bacteria, adaptation to an altered environment often includes activating transcription of virulence genes; hence, synthesis of many virulence determinants is regulated by environmental/nutritional signals (e.g., nitrogen, iron, and calcium) [12]. Because PIA is the most important virulence determinant of *S. epidermidis* [3, 11, 13-16], it is not surprising that PIA synthesis is regulated by environmental and nutritional signals [17-21].

Regulation of the *icaADBC* operon is complex, involving at least two DNA binding proteins (IcaR and SarA) and the alternative sigma factor σ^B. IcaR binds immediately 5′ to the *icaA* transcriptional start site and represses transcription of the *ica* operon [22, 23]. SarA is an essential positive effector of *icaADBC* transcription that binds to the *icaA* promoter region [24, 25]. σ^B affects PIA synthesis indirectly by regulating the expression of *icaR* [26-28]. In addition, the *luxS* quorum-sensing system has a minimal effect on PIA synthesis and biofilm formation; however, the mechanism remains to be determined [29].

In *Staphylococcus aureus*, σ^B is activated during stress conditions and growth phase transitions [30, 31]. Because *S. epidermidis icaADBC* transcription and PIA synthesis are induced by environmental and nutritional stresses, stress-dependent activation of σ^B has been a focal point of research into the environmental regulation of

PIA synthesis and biofilm formation [24, 27, 32]. Importantly, σ^B does not respond directly to environmental signals; therefore, a second area of research interest has focused on identifying the regulatory network controlling σ^B activation, which is involved in responding to environmental signals. The activation of σ^B involves an anti-sigma factor (RsbW) and an anti-anti-sigma factor (RsbV) and resembles the process in the closely related organism *Bacillus subtilis* [27]. In the absence of environmental stimuli, σ^B is bound in a complex with RsbW. Stress-inducing stimuli are hypothesized to activate the RsbU phosphatase to dephosphorylate (activate) the anti-anti-sigma factor RsbV, which then binds RsbW in a competitive manner to increase the concentration of free σ^B [26, 27]. Based on several published studies, it seems likely that the σ^B posttranslational regulatory network is responsible for responding to some environmental stresses (i.e., osmotic stress) but is only minimally involved in responding to nutritional signals (e.g., glucose), suggesting that there are other means for *S. epidermidis* to "sense" nutritional signals [19, 27, 28].

Previously, we observed the environmental and nutritional signals that enhance PIA synthesis also repress tricarboxylic acid (TCA) cycle activity, leading us to hypothesize that repression of TCA cycle activity would increase PIA accumulation [33]. To test this hypothesis, we incubated *S. epidermidis* strain 1457 with very low concentrations of fluorocitrate, which minimally inhibited TCA cycle activity, and determined the amount of cell-associated PIA. As the concentration of fluorocitrate in the culture medium was increased, the amount of PIA produced increased proportionally [33]. Fluorocitrate is a highly toxic compound when metabolized to 4-hydroxy-

*trans*aconitate by the TCA cycle enzyme aconitase, severely limiting the usable concentration range [34, 35]. In the present study, we overcame this limitation by genetically inactivating the aconitase gene in *S. epidermidis* strain 1457. The aconitase mutant and complemented strains permitted a detailed analysis of the function of central metabolism in PIA synthesis.

5.2 Materials and Methods

5.2.1 Bacterial strains, bacteriophage, plasmids, and growth conditions

Bacterial strains, bacteriophage, and plasmids used in this study are listed in Table 5.1. *S. epidermidis* strain 1457 [36] was used in these studies because it is a known PIA and biofilm-producing strain. *Escherichia coli* strains were grown in 2x YT broth or on 2x YT agar [37]. *S. aureus* and *S. epidermidis* strains were grown in tryptic soy broth without dextrose (TSB; BD Biosciences) supplemented with 0.25% glucose or on TSB containing 1.5% (wt/vol) agar. *S. epidermidis* cultures were inoculated 1:200 from overnight cultures (normalized for growth) into TSB, incubated at 37°C, and aerated at 225 rpm with a flask-to-medium ratio of 7:1. Bacterial growth was assessed by measuring the optical density at 600 nm or by determining the number of CFU mL^{-1}. Antibiotics were purchased from Fisher Scientific or Sigma Chemical and were used at the following concentrations: ampicillin (100 μg/mL for *E. coli*), chloramphenicol (8 μg/mL), erythromycin (8 μg/mL), minocycline (2 μg/mL), and trimethoprim (10 μg/mL).

Table 5.1. Bacterial strains, bacteriophage, and plasmids used in this study

Bacterial strain, bacteriophage, or plasmid	Relevant properties[a]	Source or reference
Bacterial Strains		
DH5α	*E.coli* cloning host	Invitrogen
RN4220	*S. aureus* restriction-deficient strain 8325-4	[38]
1457	*S. epidermidis* biofilm-forming strain	[36]
1457-*acnA*	Strain 1457 *acnA* inactivated mutant; Min^r	This study
1457-*citZC*	Strain 1457 *citCZ* deletion mutant; Erm^r	This study
1457-*Δica*	Strain 1457 *icaADBC* deletion mutant; Tmp^r	[24]
1457-*icaR*	Strain 1457 *icaR* inactivated mutant; Min^r	[24]
1457-*sigB*	Strain 1457 *sigB* inactivated mutant; Tmp^r	[24]
1457-*sarA*	Strain 1457 *sarA* inactivated mutant; Min^r	[24]
Bacteriophage 71	*S. epidermidis* transducing phage	[39]
Plasmids		
pCL15	Expression vector; derivative of pSI-1; Cam^r	Chia Lee
pEC4	pBluescript II KS(+) with *ermB* inserted into ClaI site; Amp^r Erm^r	[40]
pJF12	Plasmid pCR2.1 containing *tetM*; Amp^r Min^r	J. Finan and G. Archer
pGEM-T	*E. coli* TA cloning vector; Amp^r	Promega
pMRS2	pCL15 with *acnA* gene under the control of the P_{spac} promoter; Cam^r	This study
pMRS5	Derivative of pTS1-d with *citZC::ermB* fragment; Amp^r Cam^r Erm^r	This study
pNF103	pUC19 with 1.7-kb internal fragment of *acnA*; Amp^r	This study
pNF117	pNF103 with *tetM* inserted into *EcoRI* site; Amp^r Min^r	This study
pNF118	pRO6448::pNF117; Amp^r Erm^r Min^r	This study
pRO6448	pE194(ts) with AluI fragment (*oriT*) of pC221 cloned into the unique ClaI site; Erm^r	[41]
pTS1	Shuttle vector; pE194*ori*^{ts}; ColE1; Amp^r Cam^r	[42]
pTS1-d	Derivative of pTS1 with deletion of plasmid-encoded 3′ region of ermC	This study
pUC19	*E. coli* cloning vector; Amp^r	[43]

[a] Amp^r, ampicillin resistant; Cam^r, chloramphenicol resistant; Ermr, erythromycin resistant; Min^r, minocycline resistant; Tmp^r, trimethoprim resistant; ts, temperature-sensitive origin of replication.

Table 5.2. Oligonucleotide primers used in this study

Primer	Nucleotide sequences (5′→ 3′)
663 forward	GCGCGCGGCGCCGGGATATAAAATGGCTTCTA
664 reverse	GCGCGCCTGCAGACCAGCAAGCGCATAAGCTA
SERP1762(glmM)-f	CACCTGAATTGGCTTTTAAACTAGG
SERP1762(glmM)-r	TCAACACGATATTTGTCAGTCACTC
SERP1760(glmS)-f	GTGTGGAATTGTTGGTTATATTGGC
SERP1760(glmS)-r	GCTTCACGTGATAATGTTGAACC
glmU-f	CGATTATTCTGGCAGCAGGTAAG
glmU-r	CGATTGATACGTTGTTGCAAAGC
icaProbefoward	GACAGTCGCTACGAAAAG
icaProbereverse	CCGAATAATTTGTAAATTTCC
icaR-r	GGAGGCTCTAAAATACCAAGTCTC
pgi-f	GCATGAACTAGATCAGCAAAAGGATATTG
pgi-r	GCTGTAAGAACAGAATAACGACCTCC
sarA-f	GGCTATTTCAAAAATCAATGACTGCTTTG
sarA-r	GCTTCTGTGATACGGTTGTTTACTC
sigB-f	CGAAAGAGTCGAAATCAGCTAGTG
sigB-r	GTGAAACATGCATTTGACTAAGACC
SalI-SD-acnA-f	CTAAGTCGACAATATCAAGGGGGATATAAAA TGGCTTC
SacI-acnA-r	CAAGAGCTCATGACTG CTTGTACTCAATAGTGAG
pCL15-f	GTTCTACATCCAGAACAACCTCTGC
pCL15-r	GAAATGATGACCTCGTTTCCACCG
citC-r-ermB2	GACGTTCTGTACCTTCTTTAGGCGACTCATAGAAT TATTTCCTCCC
citZ-f-ermB	CACTTTGAGGAGTATGTGACGAAGGAGGGATTCG TCATGTTG
phoP-BamHI	GGAGGATCCACGTACAGTCTTTATCAATTTA GGTTGC
ermB-citC-f2	GGAAATAATTCTATGAGTCGCCTAAAGAAGGTA CAGAACGTCTAGTAAGAG
ansP-SacI	GCAGAGCTCACTAAGTGTCGATGTAAATTTTA TCGTGC
ermB-citZ-r	GACGAATCCCTCCTTCGTCACATACTCCTCAAA GTGCTTGTATAC

5.2.2 Staphylococcus epidermidis mutant construction. Aconitase (acnA) mutant

A 1.7-kb internal PCR product of the *acnA* gene of *S. epidermidis* strain 1457 was amplified using primers 663 forward (contains a KasI restriction site) and 664 reverse (contains a PstI restriction site) (Table 5.2) and cloned into pGEM-T (Promega). The 1.7-kb PCR product was excised from pGEM-T by digestion with KasI and PstI and ligated into the KasI and PstI sites of pUC19 [43] to generate the plasmid pNF103. The *tetM* cassette from plasmid pJF12 (from J. Finan and G. Archer) was inserted into a unique EcoRI site of pNF103 at nucleotide position 856 of the 2,702-bp *acnA* gene to generate the pNF117 plasmid. The temperature-sensitive plasmid pROJ6448 [41] was digested with PstI and ligated into the unique PstI site of pNF117. The resulting temperature sensitive plasmid pNF118 was passaged through *S. aureus* strain RN4220 and introduced into *S. epidermidis* strain 1457 by electroporation [44]. Strain 1457 containing plasmid pNF118 was used to construct an *acnA*::*tetM* mutant as described previously [45]. The resulting mutant was backcrossed into wild-type strain 1457 using transducing phage 71 [39]. Inactivation of the *acnA* gene by *tetM* cassette was verified by PCR, Southern blot analysis [46], and enzymatic assays [47].

5.2.3 Citrate synthase (citZ) and isocitrate dehydrogenase (citC) double mutant

The genes coding for citrate synthase and isocitrate dehydrogenase are present in a two-gene operon and cotranscribed. To inactivate the *citZC* operon, a gene splicing by overlap extension technique was used to replace a 2-kb internal DNA region of *citZC* by the *ermB* gene. The *ermB* gene was amplified from pEC4 [40] using primers citC-r-

ermB2 and citZ-f-ermB (Table 5.2), containing sequences homologous to the *citC* and *citZ* genes. Genomic DNA from *S. epidermidis* strain 1457 was used as a template for PCR amplification of the DNA flanking the internal *citZC* region. Primers ansP-SacI and ermB-citZ-r (Table 5.2) were used for amplification of a 1.9-kb region of the *citZ* upstream region. A 1.7-kb region of the *citC* downstream region was amplified using phoP-BamHI and ermBcitC-f2 primers (Table 5.2). The resulting 4.6-kb PCR product consisted of a 1-kb *ermB* gene with DNA flanking the internal *citZC* regions. The 4.6-kb PCR product contained SacI and BamHI sites that were used for ligation into pTS1-d digested with SacI and BamHI to generate pMRS5. The temperature-sensitive plasmid pTS1-d is a derivative of pTS1 [42] with deletion of the plasmid-encoded 3′region of the *ermC* gene. It was constructed by PCR amplification of pTS-1 DNA using primers pTS1dErm-f and pTS1dErm-r, and the resulting PCR product was self-ligated as described previously [48]. Plasmid pMRS5 was used to construct a strain 1457 *citZC* double mutant (1457-*citZC*::*ermB*) [45]. Allelic replacement of the internal region of *citZC* genes by the *ermB* cassette was verified by PCR and Southern blot analysis.

5.2.4 Construction of the complementation plasmid

Plasmid pCL15 (a kind gift of Chia Lee), containing a P_{spac} promoter, was used for the construction of an *acnA* complementation plasmid. The pMRS2 complementation plasmid was constructed by amplifying the 2.7-kb promoterless *acnA* gene from *S. epidermidis* strain 1457 using SalI-SD-acnA-f and SacI-acnA-r primers, followed by ligation of the PCR product into plasmid pCL15 digested with SalI and SacI. Plasmid

pMRS2 was isolated from *S. aureus* strain RN4220 and electroporated into *S. epidermidis* strain 1457 [44]. Plasmid pMRS2 was introduced into strain SE1457- *acnA* using transducing phage 71 [38]. Induction of expression from the P_{spac} promoter was performed by supplementing bacterial cultures with isopropyl-β-D-thiogalactopyranoside (IPTG) (1 mM).

5.2.5 Measurement of acetic acid and glucose concentrations in culture medium

Aliquots of bacteria (2 mL) were centrifuged for 3 min at 14,000 rpm at 4°C, and the supernatants were removed and stored at -20°C until use. Acetate and glucose concentrations were determined with kits purchased from R-Biopharm and used according to the manufacturer's protocol.

5.2.6 Determination of NAD^+, NADH, and ATP concentrations

Intracellular NAD^+ and NADH concentrations were determined with an enzymatic cycling assay kit (Biovision). Briefly, aliquots (5 to 20 mL) were harvested at 2, 3, 4, 5, and 6 h by centrifugation at 4°C for 10 min at 4,000 rpm. Bacterial pellets were suspended in 1 mL of extraction buffer (Biovision) and lysed using lysing matrix B tubes (Qbiogene) in a FastPrep instrument (Qbiogene). The lysate was centrifuged at 4°C for 5 min at 14,000 rpm. NAD^+ and NADH concentrations were determined according to the manufacturer's protocol and normalized to the cell density. All assays were performed in duplicate for two independent experiments.

Intracellular ATP concentrations were determined using the Enliten ATP assay (Promega). Briefly, $2x10^7$ bacteria were collected at 2, 3, 4, 5, and 6 h, washed three

times with 1 mL of phosphate-buffered saline, and suspended in 100 µL of phosphate-buffered saline. Bacteria were lysed by the addition of 900 µL of dimethyl sulfoxide [49]. ATP concentrations were determined according to the manufacturer's protocol and normalized to the cell density. All assays were performed in duplicate for two independent experiments.

5.2.7 Determination of aconitase activity

Cell-free lysates of *S. epidermidis* were prepared as follows: aliquots of bacteria (2 mL) were harvested by centrifugation and suspended in 750 µL of lysis buffer containing 90 mM Tris (pH 8.0) and 100 µM fluorocitrate. The samples were lysed using lysing matrix B tubes (Qbiogene) in a FastPrep instrument (Qbiogene). The lysate was centrifuged for 5 min at 14,000 rpm and 4°C. Aconitase activity was assayed in the resulting cell-free lysate as previously described [50]. One unit of aconitase activity is defined as the amount of enzyme necessary to give a ΔA_{240} min^{-1} of 0.0033 [51]. Protein concentrations were determined by the method of Lowry [52].

5.2.8 PIA immunoblot assay

PIA accumulation was determined as described previously [33]. Briefly, TSB medium containing 0.25% glucose was inoculated with equal numbers of bacteria from overnight cultures. The cultures were grown at 37°C, with a flask-to-medium ratio of 7:1, and aerated at 225 rpm. Every 2 h, equal numbers of bacteria (0.25 A_{600} unit) were harvested by centrifugation, and the PIA was extracted in 0.5 M EDTA (pH 8.0) by

boiling for 5 min. Aliquots of PIA were applied to a polyvinylidene difluoride membrane (Immobilon-P; Millipore) and blocked with 5% skim milk overnight. The polyvinylidene difluoride membrane was incubated for 2 h with PIA-specific antiserum (generously provided by Michael Otto), which was followed by a 2 h incubation with an antirabbit immunoglobulin G peroxidase conjugate. The presence of PIA was detected using Super Signal West Pico chemiluminescent substrate (Pierce). The integrated density values of bands on autoradiographs were determined with the TotalLab software (Nonlinear Dynamics Ltd.)

5.2.9 Northern blot analysis

Northern blot analysis of transcripts was performed as described previously [37], except that total RNA was isolated using the FastRNA Pro Blue kit (Qbiogene) and purified using an RNeasy kit (Qiagen). Oligonucleotide primers (Table 5.2) used in making DNA probes were designed using the *S. epidermidis* RP62A genome sequence (GenBank accession number CP000029). Probes for Northern blotting were generated by PCR amplification of unique internal regions of *glmM*, *glmS*, *glmU*, *icaD*, *icaR*, *pgi*, *sarA*, and *sigB* genes and labeled using the North2South random prime labeling kit (Pierce). Electrophoresis, transfer of the RNA to the Nytran SPC nylon membrane (Whatman), and hybridization were done using the NorthernMax kit (Ambion). Detection was performed using the chemiluminescent nucleic acid detection module (Pierce). When necessary, reprobing of blots was done according to the manufacturer's protocol.

5.2.10 NMR data collection, processing, and analysis

Nuclear magnetic resonance (NMR) samples were prepared from three independent cultures (50 mL) of *S. epidermidis* strains 1457 and 1457-*acnA* during the exponential growth phase (2 h) and post exponential growth phase (6 h) using [$^{13}C_6$]glucose (Cambridge Isotope Laboratories) in place of glucose in the growth medium. Bacteria were rapidly harvested, lysed using lysing matrix B tubes (Qbiogene) in a FastPrep instrument (Qbiogene), and centrifuged to remove the cell debris. Samples were lyophilized, suspended in a 99.8% D2O (Isotec) 50 mM phosphate buffer, and an aliquot (500 μL) was transferred to an NMR tube for analysis.

The NMR spectra were collected on a Bruker 500-MHz Avance spectrometer equipped with a triple-resonance, *z*-axis gradient cryoprobe. A BACS-120 sample changer with Bruker Icon software was used to automate the NMR data collection. Two-dimensional (2D) ^{1}H-^{13}C-heteronuclear single quantum coherence (^{1}H-^{13}C-HSQC) spectra were collected with a Bruker pulse sequence (HSQCETGP), solvent presaturation, and relaxation delay of 0.5 s. A total of 1,024 data points with a sweep width of 4734.85 Hz and 64 data points with a sweep width of 13834.26 Hz were collected in the ^{1}H and ^{13}C dimensions, respectively. A total of eight dummy scans and 128 scans were used to obtain each of the 2D ^{1}H-^{13}C HSQC NMR spectra. Spectra were processed using the NMRPipe software package [53] and zero-filled with 2,048 and 1,024 points in the ^{1}H and ^{13}C dimensions, respectively. Additionally, the spectra were Fourier transformed, manually phased, and baseline corrected.

Spectra were analyzed using Primitive Interactive Peak Picker (PIPP) [54] to determine chemical shifts and peak intensities. The chemical shifts were divided into three groups (Tables 5.3 and 5.4) based on the intensities of similar peaks between the wild-type and *acnA* mutant strains during the exponential and post exponential phases of growth. Peak intensities were normalized for each spectrum by dividing by the average peak intensity. The triplicate data sets were then used to calculate average peak intensities for each peak observed in the 2D ^{1}H-^{13}C HSQC spectra for the wild-type and *acnA* mutant. Similarly, a percent error was calculated for each peak by dividing the standard deviation by the average peak intensity. The average peak intensities were then used to calculate a percent difference between the wild-type and acnA mutant. Peaks with calculated percent differences greater than five times the average percent error were considered to have increased concentrations in the wild-type or acnA mutant strains. Peaks with less than a fivefold deviation were considered similar. The peaks assigned to the same metabolite were required to have the same relative change in intensity in order to be classified as a metabolite with an increase in concentration.

NMR peaks were assigned to specific metabolites using the Madison Metabolomics Consortium Database [55] using a tolerance of 0.05 ppm and 0.50 ppm for ^{1}H and ^{13}C dimensions, respectively. The presence of metabolites and metabolic pathways were verified with the KEGG [56] and Metacyc [57] databases. NMR reference data available from the Madison Metabolomics Consortium Database is incomplete, so it is currently not feasible to assign every NMR resonance observed in the 2D ^{1}H-^{13}C HSQC NMR spectra to a *S. epidermidis* metabolite. It is also important to note that the

Table 5.3. Summary of metabolic differences identified in *S. epidermidis* strains by NMR after 2 h of growth

Increased intracellular conc. in the wild-type strain 1457[a]	Increase intracellular conc. in strain 1457-acnA[a]	Similar intracellular conc. in wild-type and 1457-acnA strains
Lactic acid	Valine	Acetate
Succinate	Leucine	Glucose
Glutamate	Lysine	Glucosamine
Glutamine	Citrate	Ethanol
Asparagine	Serine	Glucose-6-phosphate
Aspartate	Arginine	Inosine
	Fructose-6-phosphate	Adenosine
	Glucono-1,5-lactone	Xanthosine
	Acetaldehyde	N-Acetylglutamate
	Isoleucine	NAD
	Ornithine	Uridine
	UDP-N-acetylglucosamine	Alanine

[a]The intracellular concentration was considered to be increased when the percent difference in the NMR peak intensities was fivefold greater than the percent error observed in the peak intensities between the triplicate NMR spectra.

Table 5.4. Summary of metabolic differences identified in *S. epidermidis* strains by NMR after 6 h of growth

Increased intracellular conc. in the wild-type strain 1457[a]	Increase intracellular conc. in strain 1457-acnA[a]	Similar intracellular conc. in wild-type and 1457-acnA strains
Glutamine	Ethanol	Glucosamine-6-phosphate
Glutamate	Alanine	Glucose
Ornithine	Lactate	Ribose
Succinate	Acetaldehyde	Glucose-6-phosphate
Asparagine	Citrate	NAD
Aspartate	UDP-N-acetylglucosamine	UDP-glucuronate
γ-Aminobutanoate	Uridine	Ribose-5-phosphate
Isocitrate		Inosine
Fructose-6-phosphate		Adenosine
		Xanthosine

[a]The intracellular concentration was considered to be increased when the percent difference in the NMR peak intensities was fivefold greater than the percent error observed in the peak intensities between the triplicate NMR spectra.

combination of 2D ^{1}H-^{13}C HSQC NMR spectra with isotopically enriched (e.g., [$^{13}C_6$]glucose) cellular extracts enables the measurement of accurate metabolite concentration changes relative to one dimensional ^{1}H NMR metabolomic methods alone. Generally, only compounds that are metabolic products of [$^{13}C_6$]glucose were detected.

5.2.11 Statistical analysis

The statistical significance of changes between wild-type and mutant strains (e.g., ATP concentrations) was assessed with the Student's *t* test. To determine whether a correlation existed between two physiological parameters, a correlation coefficient (ρ) was calculated.

5.3 Results

5.3.1 Construction and characterization of an S. epidermidis strain 1457 aconitase mutant

Partial biochemical inhibition of TCA cycle activity in *S. epidermidis* increased the accumulation of PIA [33]. Because fluorocitrate is very toxic, we were unable to completely block TCA cycle activity without killing the bacteria. To overcome this obstacle, we inactivated the TCA cycle in *S. epidermidis* strain 1457 by insertion of a *tetM* cassette into the gene encoding aconitase (*acnA*; also known as *citB*) and then assessed the effect on PIA synthesis. In *S. epidermidis*, aconitase is present as a sole copy on the chromosome; hence, inactivation of aconitase results in a nonfunctional TCA cycle. Inactivation of the aconitase gene in *S. epidermidis* strain 1457 by allelic

replacement was genotypically verified by PCR and Southern blot analysis (data not shown) and phenotypically by aconitase enzymatic activity assays (Figure 5.1a). In addition, overall TCA cycle function was assessed by measuring the accumulation and depletion of acetic acid from the culture medium of the wild-type, mutant, and complemented strains (Figure 5.1d). (Acetic acid catabolism in staphylococci requires a fully functional TCA cycle [47, 57].) As expected, aconitase inactivation inhibited the post exponential oxidative catabolism of acetic acid (Figure 5.1d). Of importance, when the concentration of acetic acid in the culture medium was plotted as a function of growth, the exponential growth phase rates of acetic acid accumulation in strains 1457 and 1457-*acnA* were equivalent (data not shown), suggesting that carbon utilization in both strains was proceeding through glycolysis and the acetate kinase/phosphotransacetylase pathway at a similar rate. The equivalent rate of acetic acid accumulation in strains 1457 and 1457-*acnA* was also supported by NMR metabolomic analysis (Table 5.3). If carbon flow through glycolysis were equivalent, then it is reasonable to predict that the growth rate of the wild-type and aconitase mutant strains would also be equivalent.

Surprisingly, the growth rate of strain 1457-*acnA* was lower than the growth rate of the isogenic wild-type strain, and this was reflected in the temporal depletion of glucose (Figure 5.1c). This result was surprising because TCA cycle activity is repressed during the exponential growth phase when nutrients are abundant; thus, inactivating a repressed metabolic pathway should have minimal consequences. Restoration of the growth rate and the ability to catabolize acetate was accomplished by complementation of

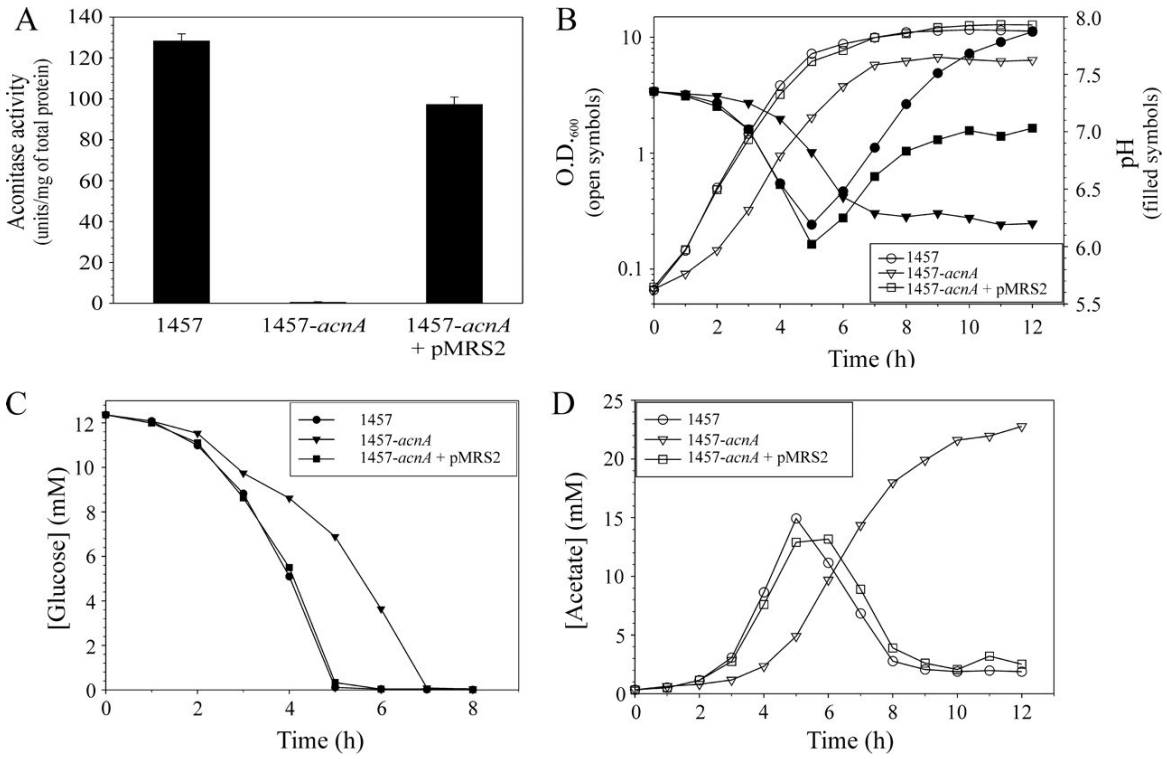


Figure 5.1. Growth characteristics of the *S. epidermidis* strain 1457 aconitase mutant. **a)** Aconitase activity of *S. epidermidis* strains 1457, 1457-*acnA*, and 1457-*acnA* containing pMRS2 (1457-*acnA* _ pMRS2) was determined 6 h postinoculation. **b)** Growth curves of wild-type strain 1457, aconitase mutant strain 1457-*acnA*, and strain 1457-*acnA* containing the complementing plasmid pMRS2 grown in TSB containing 0.25% glucose. At the indicated times, aliquots were removed, and the optical density at 600 nm (O.D.600) was measured, and the pH of the culture medium was determined. **c)** Temporal depletion of glucose from the culture medium of strains 1457, 1457-*acnA*, and 1457-*acnA* containing pMRS2. **d)** Temporal accumulation and depletion of acetic acid in the culture medium of strains 1457, 1457-*acnA*, and 1457-*acnA* containing pMRS2. The results are representative of at least two independent experiments.

the strain 1457-*acnA* mutation with a plasmid (pMRS2) containing the *acnA* gene under the control of a P_{spac} promoter (Figure 5.1a to d). As a whole, these date demonstrate that *acnA* was inactivated, and this resulted in a metabolic block in the TCA cycle.

5.3.2 TCA cycle inactivation redirects carbon from growth into UDP-N-acetylglucosamine biosynthesis

When grown in the presence of a rapidly catabolizable carbon source, gram-positive bacteria repress the TCA cycle; hence, there is an accumulation of acidic secondary metabolites (e.g., acetate) in the culture medium. Derepression of the TCA cycle occurs upon depletion of the readily catabolizable carbon source(s) and coincides with the depletion of acetate from the culture medium. The processing of glucose through glycolysis and the resulting accumulation of acetate during substrate-level phosphorylation were equivalent in *S. epidermidis* strains 1457 and 1457-*acnA*; hence, the lower growth rate of strain 1457-*acnA* relative to the wild-type strain was likely due to a redirection of carbon and energy from growth into other cellular processes or metabolic pathways. If carbon and/or energy were diverted from growth to another metabolic pathway or process, then the number of bacteria produced per unit of glucose consumed should be lower in the aconitase mutant strain than in the wild-type strain. Plotting the glucose concentrations in the culture medium as a function of optical density at 600 nm for three independent experiments (Figure 5.2) confirmed that fewer bacteria are produced per unit of glucose consumed by strain 1457-*acnA* than by strain 1457. These data strongly suggest that carbon and/or energy were diverted from growth into

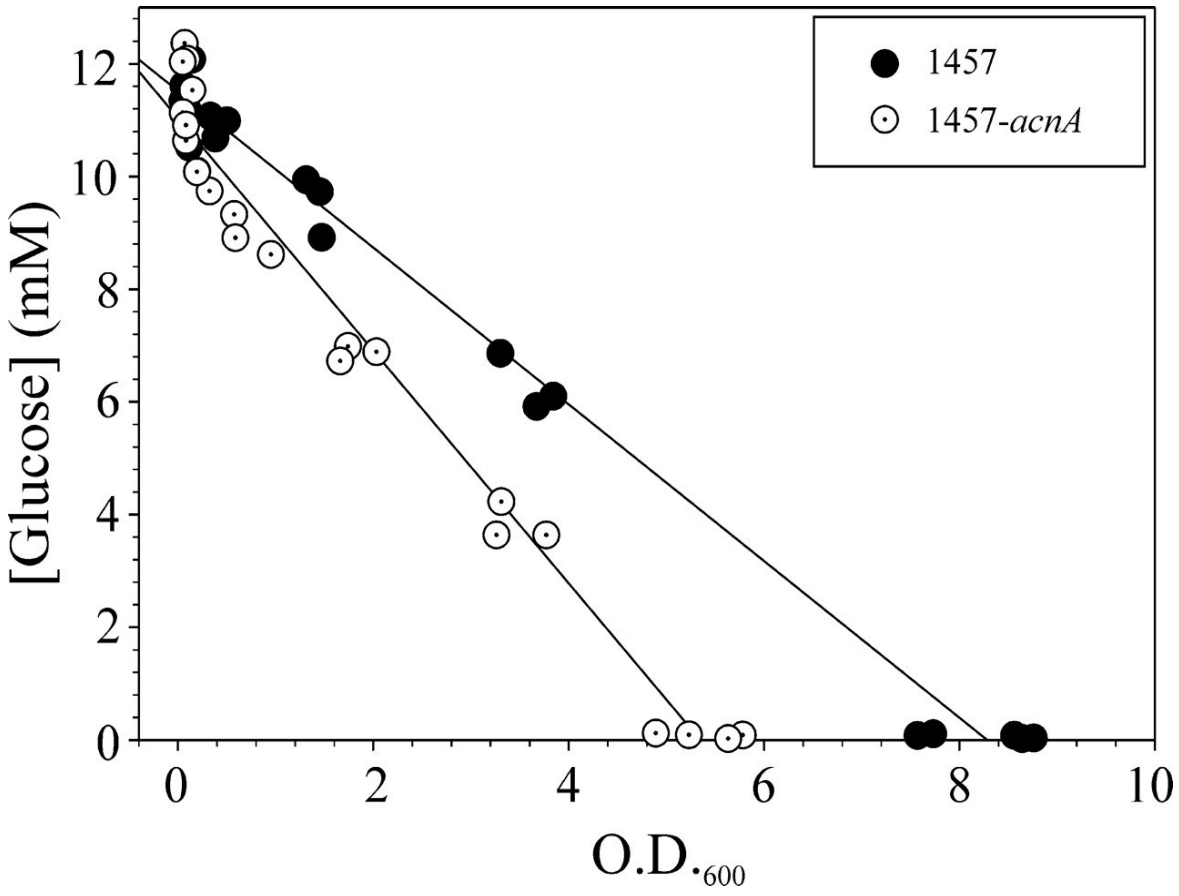


Figure 5.2. TCA cycle inactivation redirects glucose carbons away from growth. The concentrations of glucose in the culture medium of *S. epidermidis* strains 1457 and 1457-*acnA* were plotted as a function of growth. The data are from three independent experiments, and the line is a linear regression of these data generated using SigmaPlot 10.0. $O.D._{600}$, optical density at 600 nm.

other cellular processes. To determine the metabolic pathways into which carbon might be redirected, we utilized NMR metabolomics to assess the metabolic status of strains 1457 and 1457-*acnA* in the exponential growth phase (2 h) and the postexponential growth phase (6 h) (Figure 5.3 and 5.4) (Tables 5.3 and 5.4). As expected, TCA cycle inactivation decreased the intracellular concentration of amino acids derived from TCA cycle intermediates (Asn, Asp, Gln, and Glu) and increased accumulation of the PIA biosynthetic precursor UDP-*N*-acetylglucosamine. Interestingly, TCA cycle inactivation resulted in an exponential-growth phase accumulation of branched-chain amino acids relative to the wild-type strain (Figure 5.3) (Table 5.3). In addition to increasing the carbon flow into branched-chain amino acid biosynthesis, the presence of a high concentration of glucono-1,5-lactone in strain 1457-*acnA* suggests that inactivation of the TCA cycle increased carbon flow into the pentose phosphate pathway.

NMR metabolomics indicated that TCA cycle inactivation redirected carbon into UDP-*N*-acetylglucosamine biosynthesis (Tables 5.3 and 5.4). If carbon were being redirected into UDP-*N*-acetylglucosamine biosynthesis, then it is likely that transcription of genes coding for the UDP-*N*-acetylglucosamine biosynthetic enzymes would be increased. To determine whether TCA cycle inactivation altered the transcription of UDP-*N*-acetylglucosamine biosynthetic genes, we probed total RNA for *pgi* (glucose-6-phosphate isomerase), *glmS* (glucosamine-6-phosphate synthase), *glmM* (phosphoglucosamine mutase), and *glmU* (UDP-*N*-acetylglucosamine pyrophosphorylase) (Figure 5.5). Consistent with metabolomic data indicating an increased intracellular concentration of fructose-6-phosphate, transcription of *pgi* was

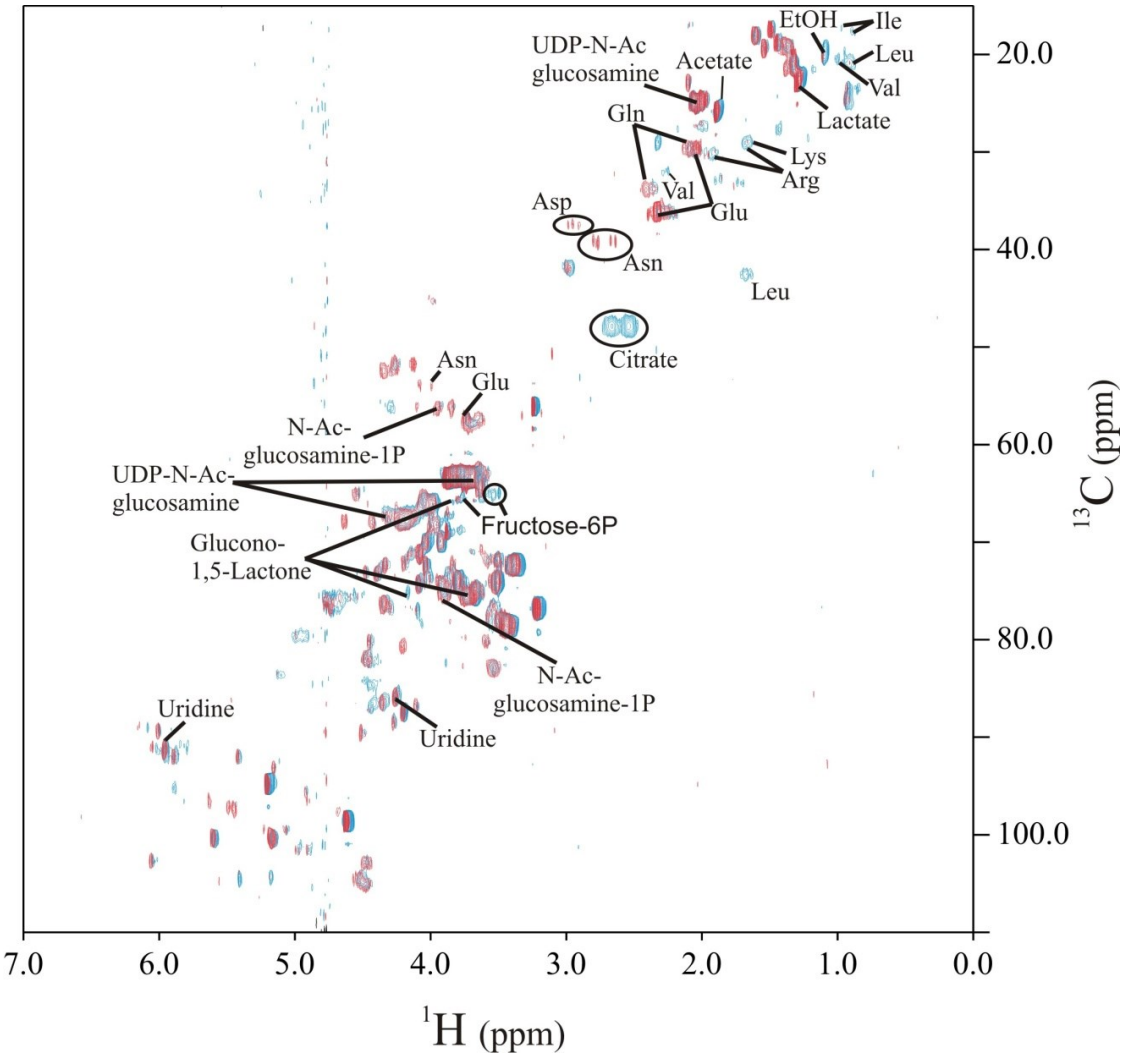


Figure 5.3. NMR overlay spectra of *S. epidermidis* strain 1457 and strain 1457-*acnA* grown for 2 hours. The spectrum for strain 1457 is displayed in red and the spectrum for strain 1475-*acnA* is displayed in blue. Peaks of metabolites of interest are indicated within the figure. The data are representative of spectra collected from three independent cultures.

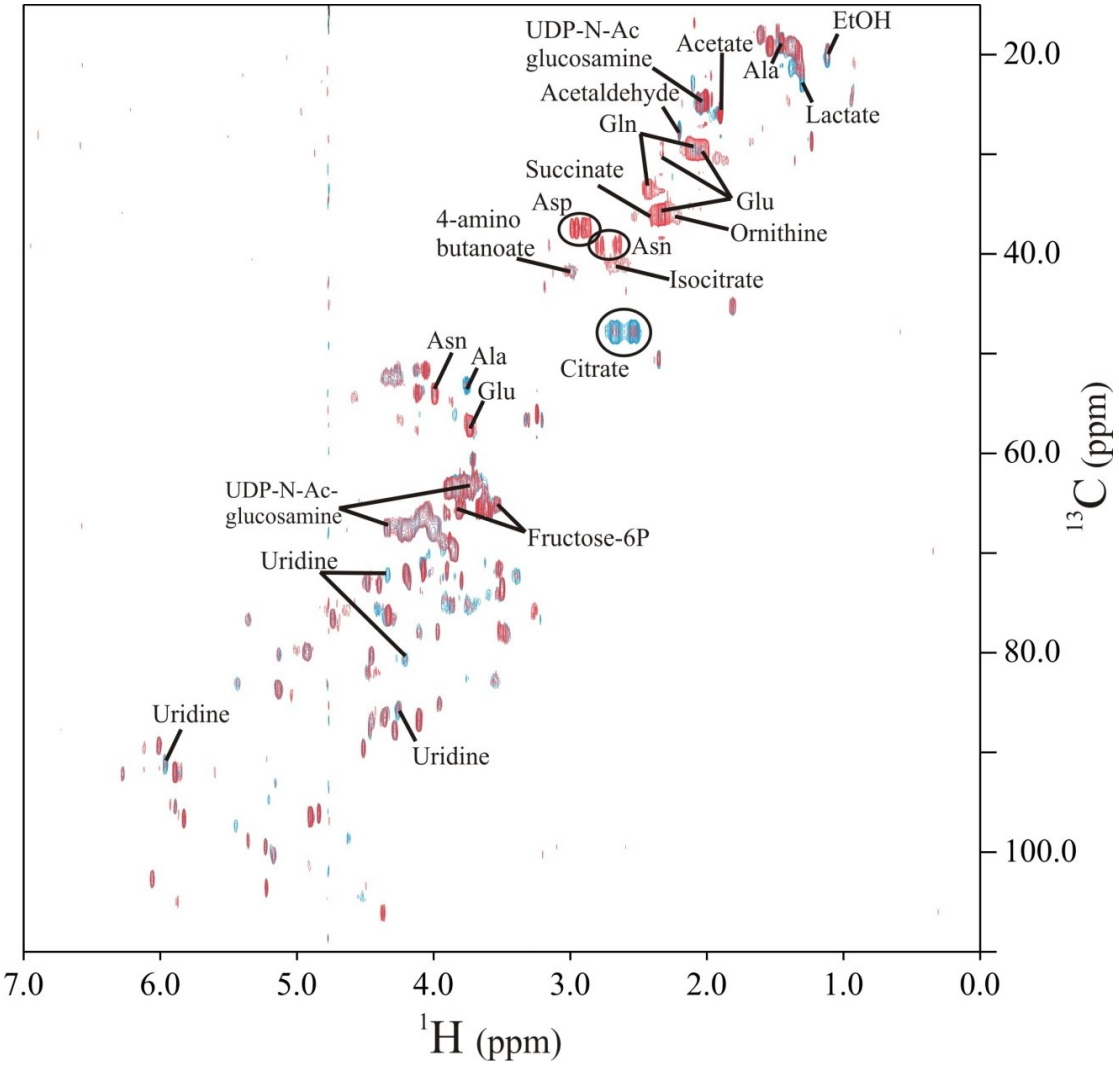


Figure 5.4. NMR overlay spectra of *S. epidermidis* strain 1457 and strain 1457-*acnA* grown for 6 hours. The spectrum for strain 1457 is displayed in red and the spectrum for strain 1475-*acnA* is displayed in blue. Peaks of metabolites of interest are indicated within the figure. The data are representative of spectra collected from three independent cultures.

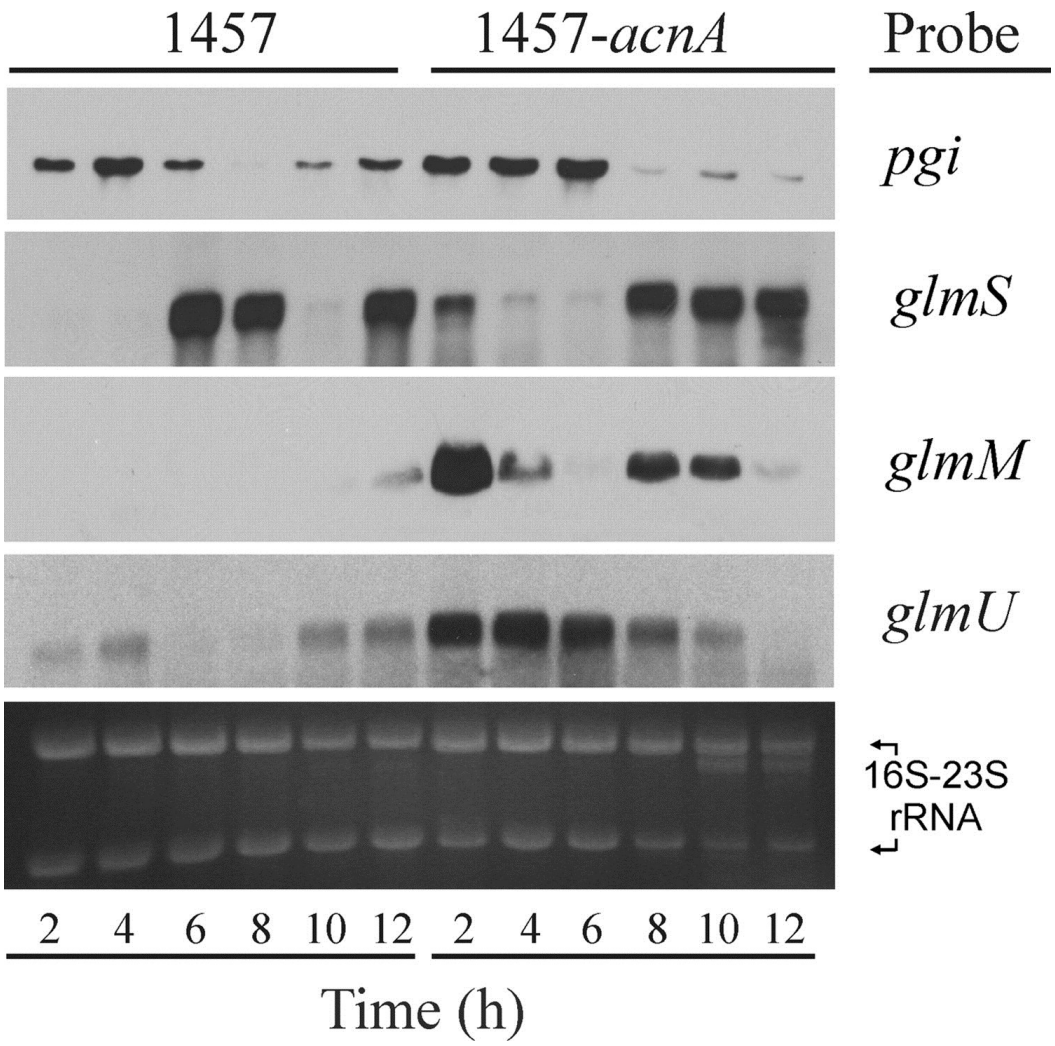


Figure 5.5. Northern blot analysis of UDP-*N*-acetylglucosamine biosynthetic genes in *S. epidermidis* strains 1457 and 1457-*acnA*. At the indicated times during the growth cycle, total RNA was isolated from strains 1457 and 1457-*acnA*, and Northern blot analysis was performed using the indicated probes. To ensure that equivalent quantities of RNA were loaded, 23S and 16S rRNA were visualized by ethidium bromide staining and used as loading controls (bottom panel). The results are representative of at least two independent experiments.

slightly increased in *S. epidermidis* strain 1457-*acnA* relative to strain 1457 during the exponential growth phase. (Glucose-6-phosphate isomerase catalyzes the second step in glycolysis, specifically, the reversible isomerization of glucose-6-phosphate into fructose-6-phosphate.) Fructose-6-phosphate is the biosynthetic precursor of UDP-*N*-acetylglucosamine, whose biosynthesis is encoded within the genes *glmS*, *glmM*, and *glmU*. The exponential-growth-phase mRNA levels of *glmS*, *glmM*, and *glmU* were increased in strain 1457-*acnA* relative to strain 1457. Taken together, these data demonstrate that TCA cycle inactivation results in transcriptional changes that redirect carbon into *N*-acetylglucosamine biosynthesis.

In addition to providing biosynthetic intermediates, the TCA cycle generates ATP and reducing potential; therefore, a metabolic block in the TCA cycle will alter the energy status of the bacterium. As an indicator of the energy status, the intracellular concentrations of NAD^+, NADH, and ATP were determined for *S. epidermidis* strains 1457 and 1457-*acnA*. The intracellular concentration of ATP was lower in the aconitase mutant strain relative to the wild-type strain during both the exponential and post exponential growth phases (Figure 5.6c). This decreased ATP concentration was likely due to decreased oxidative phosphorylation, as the intracellular concentrations of NAD^+ and NADH were significantly ($P < 0.01$) lower in the aconitase mutant strain than in the wild-type strain (Figure 5.6a and b).

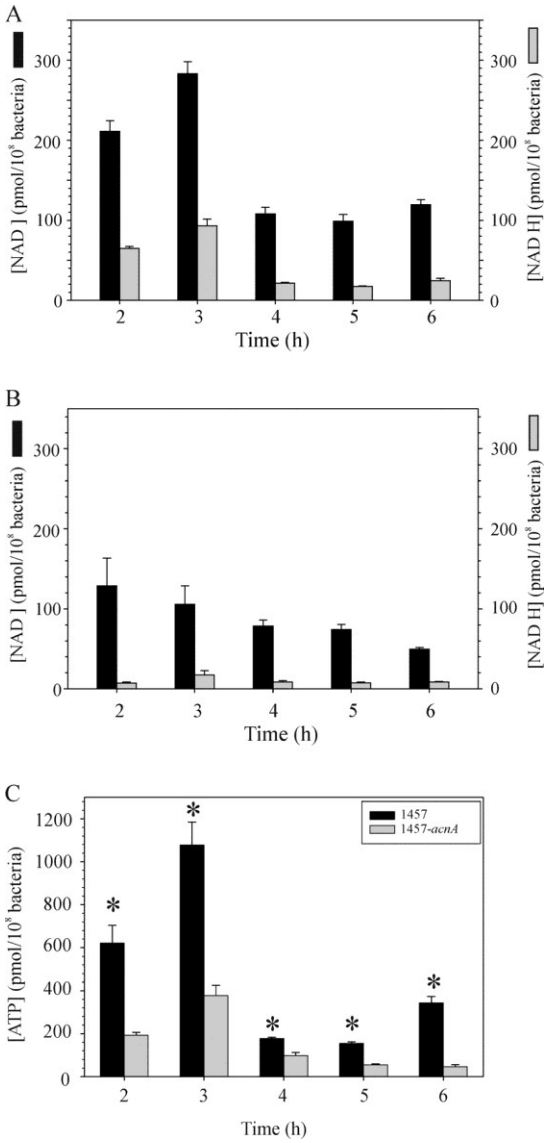


Figure 5.6. Aconitase inactivation alters the redox and energy status of *S. epidermidis*. **a)** At the indicated times during the growth cycle, intracellular concentrations of NAD and NADH were determined in *S. epidermidis* strain 1457. **b)** Similarly, the concentrations of reduced and oxidized NAD were determined for strain 1457-*acnA*. **c)** ATP concentrations were determined for strains 1457 and 1457-*acnA*. The results are presented as the means plus standard errors of the means (error bars) of duplicate determinations for at least two independent experiments.

5.3.3 TCA cycle inactivation increases PIA accumulation

To determine whether TCA cycle inactivation resulted in increased synthesis of PIA, we assessed the relative amounts of cell associated PIA in the *acnA* mutant, the complemented *acnA* mutant, and the wild-type strains using a PIA immunoblot assay (Figure 5.7a and b). In addition, to facilitate comparison of the effect of TCA cycle inactivation on PIA accumulation with the inactivation of well-established regulators (i.e., IcaR, σ^B, and SarA), we determined the relative amounts of cell-associated PIA for isogenic strains that had *icaR*, *sarA*, or *sigB* genetically inactivated. In agreement with previous studies [22, 24, 25, 27, 32], we observed that mutations in the genes of positive regulators *sigB* and *sarA* decreased PIA accumulation relative to the wild-type strain, while inactivation of *icaR* increased PIA accumulation (Figure 5.7a and b). Surprisingly, aconitase inactivation had a much greater effect on PIA synthesis than did inactivation of the well-studied repressor IcaR (Figure 5.7a and b). Complementation of the *acnA* mutation restores PIA accumulation to that seen in the parental strain (Figure 5.7a and b).

In *B*. *subtilis*, aconitase inactivation results in the accumulation of citric acid, which is partially responsible for blocking sporulation and causing a slight decrease in the growth rate [58]. NMR metabolomics revealed an accumulation of citric acid in the aconitase mutant strain relative to the wild-type strain (Figure 5.3 and 5.4) (Tables 5.3 and 5.4), raising the possibility that the increased accumulation of PIA and the decreased growth rate of *S. epidermidis* strain 1457-*acnA* relative to strain 1457 might be due to citric acid-induced stress. To address these possibilities, we inactivated both citrate synthase (*citZ*) and isocitrate dehydrogenase (*citC*) in *S. epidermidis* strain 1457 and

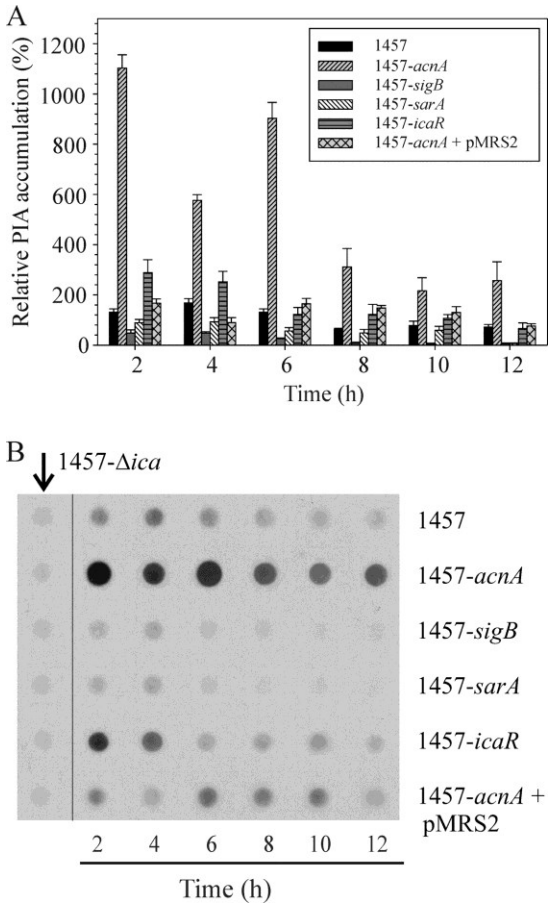


Figure 5.7. TCA cycle inactivation increases PIA accumulation. **a)** Relative amounts of cell-associated PIA were determined using a PIA immunoblot assay for *S. epidermidis* strains 1457, 1457-*acnA*, 1457-*sigB*, 1457-*sarA*, 1457-*icaR*, and 1457-*acnA* containing pMRS2 (1457-*acnA* + pMRS2). To determine the relative amounts of PIA, we defined the amount of PIA produced by strain 1457 at 2 h in one experiment as 100% and expressed all other amounts as relative to the amount at this 2-h point. The results are presented as the means plus standard errors of the means (error bars) for three independent experiments. **b)** Representative immunoblot used for determining spot volumes. Strain 1457-Δ*ica*, which does not synthesize PIA, was used to determine the background density.

assessed growth and assayed for PIA accumulation (data not shown). Similar to the aconitase mutant strain, the *citCZ* double mutant dramatically increased PIA accumulation. Additionally, the *citCZ* double mutant had a decreased growth rate relative to the wild-type strain, but the growth rate was increased relative to strain 1457-*acnA* (data not shown), suggesting that citric acid accumulation was partially responsible for the slower growth of the aconitase mutant. Taken together, these data demonstrate that TCA cycle activity represses *S. epidermidis* PIA accumulation and that this effect is independent of citric acid.

5.3.4 TCA cycle activity represses icaADBC transcription

TCA cycle inactivation alters the metabolic status of *S. epidermidis* (Figure 5.1a to d and 4.5 and Tables 5.3 and 5.4), raising the possibility that an *icaADBC* transcriptional regulator could respond to the altered metabolic environment to increase *icaADBC* transcription. To determine whether the increased accumulation of PIA by the *acnA* mutant strain was due to altered *icaADBC* transcription, we examined *icaADBC* mRNA levels in *S. epidermidis* strains 1457 and 1457-*acnA*. In addition, *icaADBC* mRNA levels were determined for strain 1457 with mutations in *sigB*, *sarA*, or *icaR*. Consistent with the PIA immunoblot data (Figure 5.7a and b), aconitase inactivation resulted in a very large increase in *icaADBC* transcription, or mRNA stability, relative to the isogenic wild-type strain (Figure 5.8a and b), demonstrating that TCA cycle activity represses *icaADBC* transcription. In agreement with previous results, we observed that *sigB* and *sarA* inactivation repressed *icaADBC* transcription and that *icaR* inactivation

derepressed *icaADBC* transcription [22, 24, 59]. Interestingly, the extent of derepression in the 1457-*acnA* mutant strain was dramatically higher than seen in an isogenic *icaR* mutant (Figure 5.8a and b), suggesting TCA cycle activity has a greater role in regulating *icaADBC* transcription than does IcaR. These data demonstrate that, like regulation of the *lac* operon, *icaADBC* transcription is regulated by a catabolic process, specifically, the TCA cycle.

5.3.5 TCA cycle inactivation alters transcription of icaR, sarA, and sigB

As stated, transcription of *icaADBC* requires at least two positive effectors, SarA and σ^B, and one negative regulator, IcaR. The dramatic increase in *icaADBC* mRNA levels in the aconitase mutant strain relative to the wild-type strain could be due to an increased number of positive effectors or a decreased number of negative effectors. To determine whether aconitase inactivation altered transcription of the known regulators of *icaADBC* transcription, we probed total RNA isolated throughout the growth cycle for changes in the transcription of *sigB*, *sarA*, and *icaR* (Figure 5.9).

Surprisingly, aconitase inactivation increased transcription or mRNA stability of both positive regulators *sigB* and *sarA* and the negative regulator *icaR*. These data suggest that TCA cycle inactivation can alter the mRNA levels of the known regulators of PIA synthesis; however, they fail to establish that these regulators are involved in TCA cyclemediated regulation of PIA synthesis. To determine whether IcaR, SarA, or σ^B is regulating PIA synthesis in response to TCA cycle-associated metabolic changes, *icaR*, *sarA*, and *sigB* regulatory mutants of strain 1457 were incubated with 400 μM of the

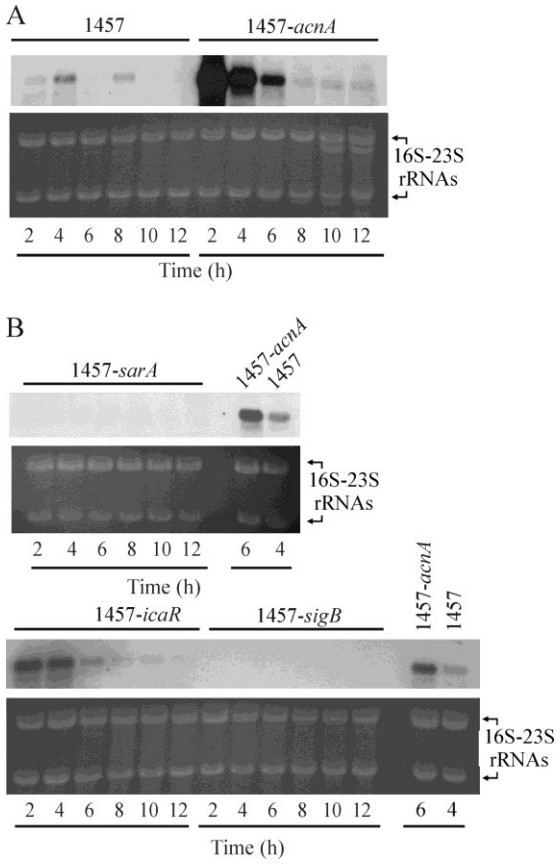


Figure 5.8. TCA cycle inactivation derepresses *icaADBC* transcription. Total RNA was isolated from *S. epidermidis* strains 1457 and 1457-*acnA*, 1457-s*arA*, 1457-*icaR*, and 1457-*sigB*, and Northern blot analysis was performed using an *icaD*-specific probe. To make certain equivalent quantities of RNA were loaded, 23S and 16S rRNA were visualized by ethidium bromide staining and used as loading controls (bottom blots). To permit direct comparison of autoradiographs, total RNA from strains 1457-*acnA* and 1457 isolated at 6 h and 4 h postinoculation, respectively, were used to normalize the exposure times. (To minimize variation, the same preparations of RNA were used as exposure controls in all experiments.) The results are representative of at least two independent experiments.

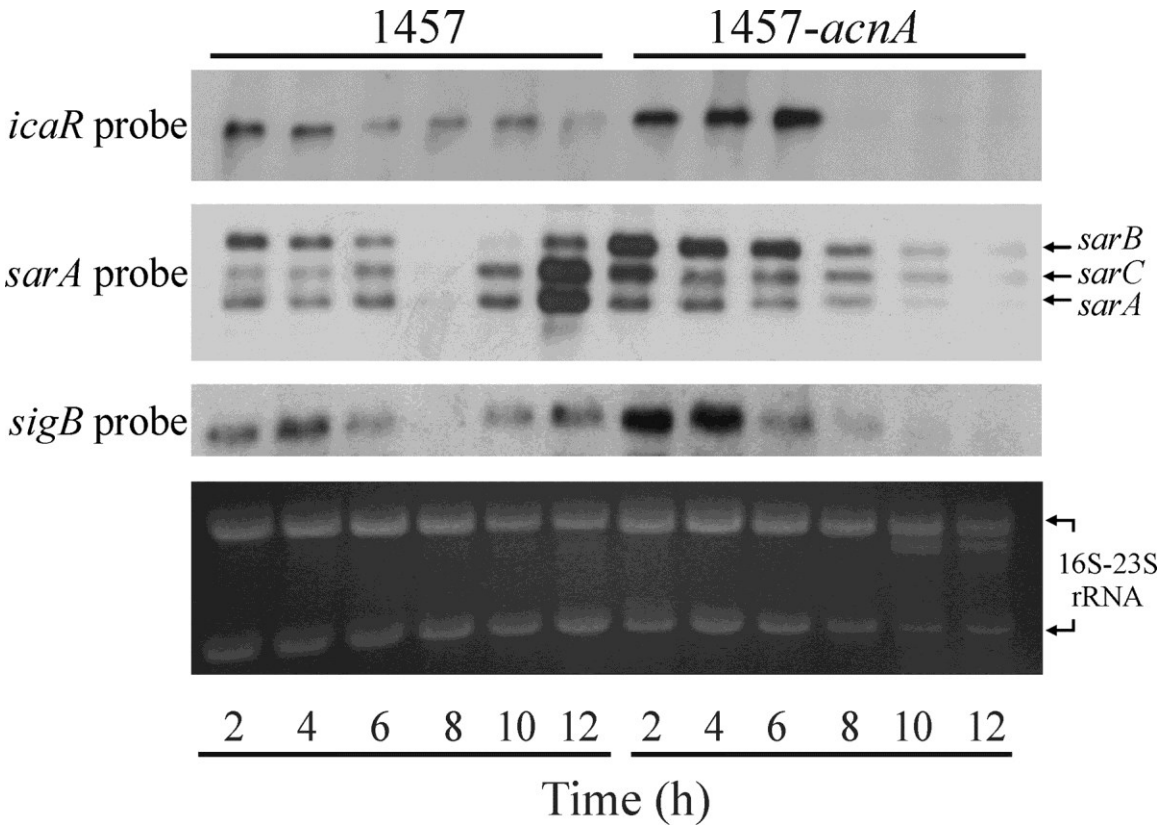


Figure 5.9. Aconitase inactivation increases transcription of *icaR*, *sarA*, and *sigB*. Northern blot analysis with *icaR*, *sarA*, and *sigB* with DNA probes generated using primers shown in Table 5.2. To ensure that equivalent quantities of RNA were loaded, 23S and 16S rRNA were visualized by ethidium bromide staining and used as loading controls (bottom blot). The results are representative of at least two independent experiments.

aconitase inhibitor fluorocitric acid, and PIA accumulation was assessed relative to the parental strain 1457 (Figure 5.10) The addition of fluorocitric acid to the culture medium of *S. epidermidis* strains 1457-*sigB* and 1457-*icaR* significantly ($P < 0.01$) increased the accumulation of PIA relative to the untreated control cultures. In strain 1457-*sarA*, the accumulation of PIA was similar irrespective of the presence of fluorocitric acid in the culture medium ($P > 0.05$). Overall, these data suggest that TCA cycle-mediated derepression of *ica* transcription and PIA synthesis occur independently of IcaR and σB. Furthermore, these data confirm that PIA synthesis requires SarA (10, 74) and at least one additional TCA cycle responsive negative regulator.

5.4 Discussion

5.4.1 Metabolic effect of TCA cycle activity on PIA biosynthesis

Glycolysis is the catabolic pathway that converts glucose into pyruvate, with the catabolic fate of pyruvate being determined by the availability of oxygen. Interestingly, the rate of glucose utilization by the wild-type *S. epidermidis* strain 1457 appears greater than the rate by strain 1457-*acnA* (Figure 5.1c); however, these data fail to account for the number of bacteria in the medium. When this is taken into account, glucose utilization was greater in strain 1457-*acnA* than in strain 1457 (Figure 5.2 c), suggesting that carbon flow through glycolysis might be greater in a TCA cycle mutant. If the amount of carbon flowing through glycolysis were greater in a TCA cycle mutant than in the parental strain, then the amount of acetate in the culture medium at the time when glucose was depleted should be greater in the mutant. Glucose was depleted in the wild-type and

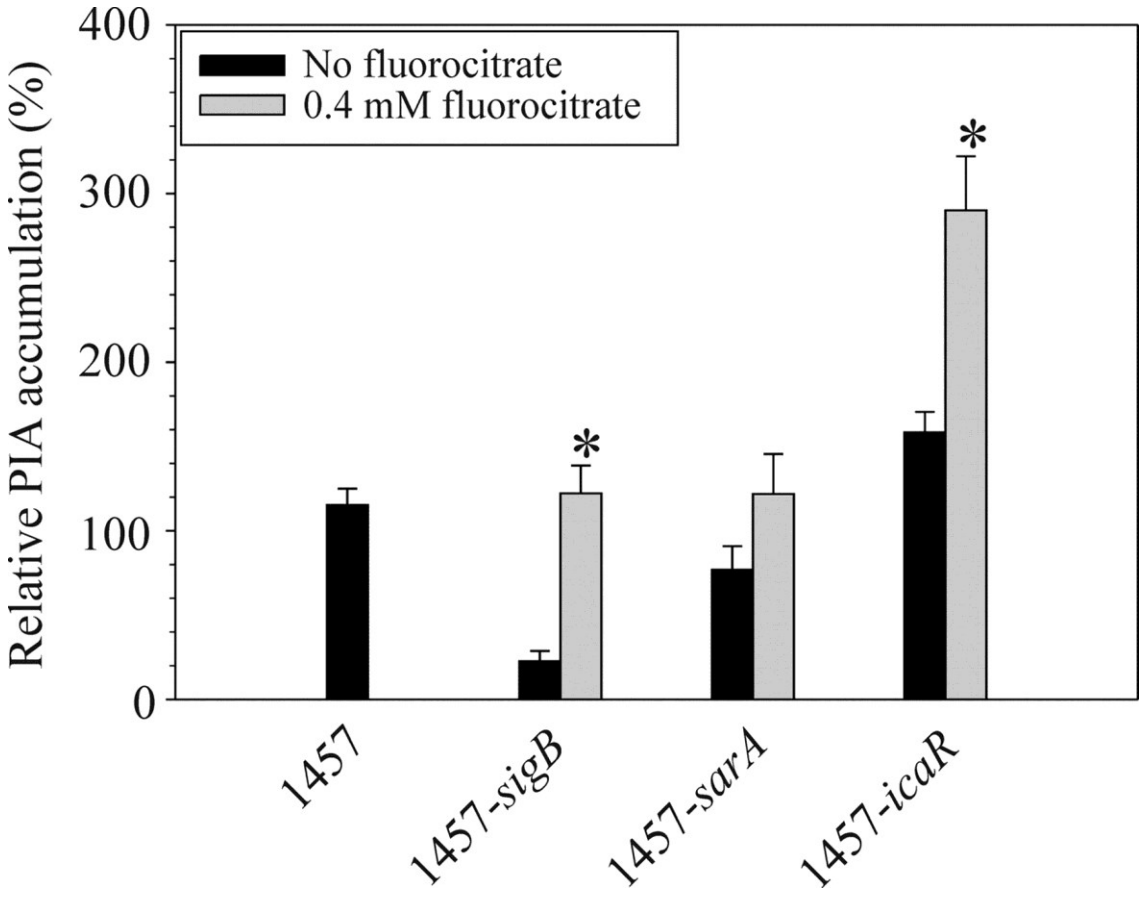


Figure 5.10. Partial biochemical inhibition of aconitase increases PIA synthesis in *sigB* and *icaR* mutants. Bacterial cultures were incubated with 400 μM fluorocitric acid and grown for 2 h, and the relative amounts of cell-associated PIA were determined by a PIA immunoblot assay. The data are presented as the means plus standard errors of the means (error bars) from three independent experiments. Values that were significantly different ($P < 0.05$) from the value for the isogenic untreated control (0 μM fluorocitrate) are indicated by an asterisk.

complemented strains after 5 h of growth; however, in strain 1457-*acnA*, glucose was depleted after 7 h of growth (Figure 5.1c). Upon depletion of glucose, the concentrations of acetate in the culture medium were similar (14 to 16 mM) for all strains (Figure 5.1d), suggesting that carbon flow through glycolysis was similar in the aconitase mutant and the wild-type strain. Additionally, because the concentrations of acetate were below 24 mM (if all glucose is processed through the glycolytic pathway, then a 2:1 stoichiometry of acetate to glucose should be maintained), the data confirm that carbon are flowing into other pathways. Consistent with these data, NMR metabolomics indicated that the exponential-growth-phase concentrations of acetic acid in strains 1457 and 1457-*acnA* were similar (Table 5.3). In addition to glycolysis, another major pathway staphylococci use for glucose catabolism is the pentose phosphate pathway [60, 61]. The presence of glucono-1,5- lactone in the NMR metabolomic data (Table 5.3) suggests that carbon flow through the pentose phosphate pathway is increased in the TCA cycle mutant strain relative to the wild-type strain. One function of the pentose phosphate pathway is to provide the five-carbon intermediate for purine biosynthesis; hence, the increased carbon flow through the pentose phosphate pathway is likely linked to the decreased intracellular concentration of ATP in strain 1457-*acnA* relative to strain 1457 (Figure 5.5). (De novo biosynthesis of ATP requires 5-phosphoribosyl-1-pyrophosphate. An important function of the pentose phosphate pathway is to provide biosynthetic intermediates for purine biosynthesis, namely, ribose-5-phosphate that is converted into 5-phosphoribosyl-1-pyrophosphate.)

NMR metabolomic data revealed an increased intracellular concentration of fructose-6-phosphate in the TCA cycle mutant relative to the wild-type strain (Table 5.3). Fructose-6-phosphate can undergo a transamination reaction to generate glucosamine-6-phosphate, a reaction catalyzed by glucosamine-6-phosphate synthase (*glmS*; EC 2.6.1.16). Phosphoglucosamine mutase (*glmM*; EC 5.4.2.10) catalyzes the reversible isomerization glucosamine-6-phosphate to glucosamine-1-phosphate. Glucosamine-1-phophaste is the substrate for the bifunctional enzyme UDP-*N*-acetylglucosamine pyrophosphorylase (*glmU*; EC 2.7.7.23 and EC 2.3.1.157) that catalyzes the acetylation and UTP-dependent activation of glucosamine-1-phosphate to generate UDP-*N*-acetylglucosamine. NMR metabolomics confirmed that *S. epidermidis* strain 1457-*acnA* has a greater intracellular concentration of UDP-*N*-acetylglucosamine than strain 1457 does. In addition, transcription of *glmS*, *glmM*, and *glmU* is increased more during the exponential growth phase in the TCA cycle mutant than in the wildtype strain (Figure 5.5). In total, these data demonstrate that TCA cycle inactivation increased the availability of the PIA monomer, UDP-activated *N*-acetylglucosamine and that this was partially due to transcriptional derepression of *N*-acetylglucosamine biosynthetic genes.

In addition to PIA synthesis, UDP-*N*-acetylglucosamine is critical for biosynthesis of the cell wall components teichoic acid and peptidoglycan. Inactivation of aconitase in *S. epidermidis* decreased the growth rate (Figure 5.1b) and the number of bacteria produced per unit of glucose (Figure 5.2), suggesting that UDP-*N*-acetylglucosamine was redirected from cell wall biosynthesis to PIA synthesis (Figure 5.7a). Thus, the decreased growth rate of the aconitase mutant strain was likely due to a combination of increased

competition for substrate (i.e., UDP-*N*-acetylglucosamine) by PIA biosynthetic enzymes and an accumulation of citric acid.

5.4.2 Regulatory effect of TCA cycle activity on PIA biosynthesis

Two common mechanisms by which bacteria control carbon flow are feedback inhibition of enzyme activity and regulation of enzyme synthesis. Undoubtedly, some of the metabolic changes arising from TCA cycle inactivation are the result of feedback inhibition of enzyme activity; however, by definition, the derepression of *icaADBC* transcription (Figure 5.8a and b) must be at the regulatory level. Although the prospect that TCA cycle enzymes directly regulate *icaADBC* transcription is a possibility, it is more likely that regulation is being mediated by regulatory proteins responding to changes in the metabolic status of the bacteria. Metabolic regulation was one of the first bacterial regulatory modalities identified, resulting in thousands of studies addressing all aspects of regulation. This long history of research into metabolic regulation has identified a large number of regulatory proteins capable of responding to changes in the intracellular concentrations of molecules associated with TCA cycle activity, such as NADH (e.g., Rex [62], NmrA [63], and CcpA [64]), citric acid (CcpC [65, 66]), glutamate (GadX and GadW [67]), branched-chain amino acids (CodY [68]), and ATP (KinA [69]). These observations lead us to hypothesize that a regulatory protein or proteins respond to TCA cycle-associated metabolites and regulate *icaADBC* transcription and PIA synthesis.

TCA cycle inactivation increased the transcription or mRNA stability of the known regulators of *icaADBC* transcription, specifically, *sigB*, *sarA*, and *icaR* (Figure 5.9). These data raise the possibility that TCA cycle-induced alterations in the metabolic status of the bacterium might be "sensed" by σ^B, SarA, or IcaR. Incubation of *S. epidermidis* strain 1457-*sigB* with a low concentration of the TCA cycle inhibitor fluorocitric acid restored PIA synthesis to wild-type levels (Figure 5.10). This fluorocitrate- induced derepression of PIA synthesis in the *sigB* mutant strain is similar to that observed in an *S. epidermidis rsbU* mutant strain (RsbU is an activator of σ^B) treated with a PIA synthesis-stimulating concentration of ethanol [26]. In that study, the authors found that ethanol could restore cell associated PIA accumulation in an *rsbU* mutant strain to a level equivalent to that of the nonstimulated wild-type strain. Because ethanol is an inhibitor of TCA cycle activity [70, 71], we hypothesize that the PIA synthesis-stimulating properties of ethanol may be due to its TCA cycle inhibitory properties.

Inactivation of *sarA* in *S. epidermidis* dramatically decreases *icaADBC* transcription [25] and PIA accumulation [24, 25] and results in a biofilm-negative phenotype [32]. Because SarA is thought to be essential for PIA synthesis [25], the more likely outcome of incubating *S. epidermidis* strain 1457-*sarA* with fluorocitric acid would be equivalent PIA accumulation between the wild-type and *sarA* mutant strains. As expected, the amount of PIA produced by strain 1457-*sarA* was similar irrespective of the presence of fluorocitric acid (Fig. 8). These data are consistent with SarA being essential for PIA synthesis and/or being required for responding to TCA cycle-associated signals.

TCA cycle inactivation increased transcription or mRNA stability of *icaR* relative to the wild-type strain (Figure 5.9). IcaR is a transcriptional repressor of *icaADBC* that was originally hypothesized to be involved in environmental regulation of PIA synthesis [22]. For these reasons, IcaR appeared to be a possible candidate for responding to TCA cycle-induced metabolic changes. If IcaR were responding to TCA cycle-associated metabolic changes, then in an *icaR* mutant background, the synthesis of PIA should be unaffected by the presence of fluorocitric acid. Growth of strain 1457-*icaR* in the presence of a low concentration of fluorocitric acid significantly increased the accumulation of PIA (Figure 5.10), demonstrating that a second repressor is involved in regulating PIA synthesis. It is likely that this second repressor is responding to TCA cycle-associated metabolic signals.

NMR metabolomics (Table 5.3) indicated that the intracellular concentrations of branched-chain amino acids were greater in the aconitase mutant strain than in the wild-type strain. In gram-positive bacteria, the intracellular concentrations of branched-chain amino acids control the expression of numerous genes, a mechanism requiring the CodY regulatory protein [68, 72, 73]. The affinity of CodY for its operator site is enhanced in response to high intracellular concentrations of GTP and branched-chain amino acids [68, 74]. The more likely explanations for our data are that CodY is an activator of *icaADBC* transcription, CodY-mediated repression is not functioning in *S. epidermidis* during the exponential growth phase, or a positive effector, potentially SarA, can overcome CodY mediated repression. We recently demonstrated that CodY is a repressor of PIA synthesis in the closely related species *S. aureus* [48]; thus, it seems unlikely that CodY is an

activator of *icaADBC* transcription in *S. epidermidis*. Investigations are under way to determine the function of CodY in regulating *S. epidermidis* PIA synthesis.

In gram-positive bacteria, carbon catabolite repression is primarily mediated by CcpA [75]. Recently, CcpA was shown to activate transcription of *icaADBC* and increase the accumulation of PIA in *S. aureus* [76]. In *B. subtilis*, CcpA regulatory activity is controlled by interactions with phosphorylated HPr or Crh and fructose-1,6-bisphosphate or glucose-6-phosphate [77]. NMR analysis indicated that TCA cycle inactivation increased the intracellular concentration of fructose-6-phosphate (Table 5.3). Fructose-6-phosphate is the biosynthetic precursor of UDP-*N*-acetylglucosamine. In addition to its importance in PIA biosynthesis, fructose-6-phosphate can be reversibly isomerized to glucose-6-phosphate by glucose-6-phosphate isomerase, resulting in CcpA activation. Taken together, these data lead us to hypothesize that TCA cycle inactivation increases CcpA-mediated activation of *icaADBC*.

5.5 Conclusion

TCA cycle activity is repressed during the exponential growth phase; however, a basal level of activity is present. As evidenced by the data in this study, disruption of this basal activity can have pronounced metabolic consequences. TCA cycle activity is regulated by the availability of nutrients, oxygen, and iron [60, 78-81] and by certain stress-inducing stimuli, such as heat, ethanol, and antibiotics [70, 82, 83]; thus, numerous external conditions can interfere with the basal TCA cycle activity and produce metabolic changes. The linkage of TCA cycle activity and exopolysaccharide synthesis and the

susceptibility of the TCA cycle to environmental inactivation lead us to speculate that one mechanism by which staphylococci perceive external environmental change is through alterations in TCA cycle activity. Thus, we hypothesize that the TCA cycle acts as a novel signal transduction pathway to translate external stimuli/conditions into intracellular signals that can stimulate or repress the activity of regulatory proteins like CodY and CcpA.

5.6 References

1. *National Nosocomial Infections Surveillance (NNIS) System Report, Data Summary from January 1992 through June 2004, Issued October 2004.* American Journal of Infection Control, 2004. **32**(8): p. 470-85.

2. von Eiff, C., G. Peters, and C. Heilmann, *Pathogenesis of Infections Due to Coagulase-Negative Staphylococci.* The Lancet Infectious Diseases, 2002. **2**(11): p. 677-85.

3. Begun, J., et al., *Staphylococcal Biofilm Exopolysaccharide Protects Against Caenorhabditis elegans Immune Defenses.* PLoS Pathogens, 2007. **3**(4): p. e57.

4. Olson, M.E., et al., *Adherence of Staphylococcus epidermidis to Biomaterials is Augmented by PIA.* Clinical Orthopaedics and Related Research, 2006. **451**: p. 21-4.

5. Vuong, C., et al., *A Crucial Role for Exopolysaccharide Modification in Bacterial Biofilm Formation, Immune Evasion, and Virulence.* The Journal of Biological Chemistry, 2004. **279**(52): p. 54881-6.

6. Maira-Litran, T., et al., *Immunochemical Properties of the Staphylococcal Poly-N-Acetylglucosamine Surface Polysaccharide.* Infection and Immunity, 2002. **70**(8): p. 4433-40.

7. Mack, D., et al., *The Intercellular Adhesin Involved in Biofilm Accumulation of Staphylococcus epidermidis is a Linear Beta-1,6-Linked Glucosaminoglycan: Purification and Structural Analysis.* Journal of Bacteriology, 1996. **178**(1): p. 175-83.

8. Heilmann, C., et al., *Molecular Basis of Intercellular Adhesion in the Biofilm-Forming Staphylococcus epidermidis.* Molecular Microbiology, 1996. **20**(5): p. 1083-91.

9. Vuong, C., et al., *Polysaccharide Intercellular Adhesin (PIA) Protects Staphylococcus epidermidis Against Major Components of the Human Innate Immune System.* Cellular Microbiology, 2004. **6**(3): p. 269-75.

10. Rupp, M.E., *Infections of Intravascular Catheters and Vascular Devices.* In K. B. Crossley and G. L. Archer (ed.), The Staphylocci in Human Disease. Churchill-Livingstone, New York, NY, 1997: p. 379-399.

11. Rupp, M.E., et al., *Characterization of the Importance of Polysaccharide Intercellular Adhesin/hemagglutinin of Staphylococcus epidermidis in the Pathogenesis of Biomaterial-Based Infection in a Mouse Foreign Body Infection Model.* Infection and Immunity, 1999. **67**(5): p. 2627-32.

12. Mekalanos, J.J., *Environmental Signals Controlling Expression of Virulence Determinants in Bacteria.* Journal of Bacteriology, 1992. **174**(1): p. 1-7.

13. Cho, S.H., et al., *Detection of the icaADBC Gene Cluster and Biofilm Formation in Staphylococcus epidermidis Isolates from Catheter-Related Urinary Tract Infections.* International Journal of Antimicrobial Agents, 2002. **19**(6): p. 570-575.

14. Kozitskaya, S., et al., *The Bacterial Insertion Sequence Element IS256 Occurs Preferentially in Nosocomial Staphylococcus epidermidis Isolates: Association with Biofilm Formation and Resistance to Aminoglycosides.* Infection and Immunity, 2004. **72**(2): p. 1210-5.

15. Rupp, M.E., et al., *Characterization of Staphylococcus epidermidis Polysaccharide Intercellular Adhesin/hemagglutinin in the Pathogenesis of Intravascular Catheter-Associated Infection in a Rat Model.* Infection and Immunity, 1999. **67**(5): p. 2656-9.

16. Ziebuhr, W., et al., *Detection of the Intercellular Adhesion Gene Cluster (ica) and Phase Variation in Staphylococcus epidermidis Blood Culture Strains and Mucosal Isolates.* Infection and Immunity, 1997. **65**(3): p. 890-6.

17. Cramton, S.E., et al., *Anaerobic Conditions Induce Expression of Polysaccharide Intercellular Adhesin in Staphylococcus aureus and Staphylococcus epidermidis.* Infection and Immunity, 2001. **69**(6): p. 4079-85.

18. Deighton, M. and R. Borland, *Regulation of Slime Production in Staphylococcus epidermidis by Iron Limitation.* Infection and Immunity, 1993. **61**(10): p. 4473-9.

19. Dobinsky, S., et al., *Glucose-Related Dissociation Between icaADBC Transcription and Biofilm Expression by Staphylococcus epidermidis: Evidence for an Additional Factor Required for Polysaccharide Intercellular Adhesin Synthesis.* Journal of Bacteriology, 2003. **185**(9): p. 2879-86.

20. Gotz, F., *Staphylococcus and Biofilms.* Molecular Microbiology, 2002. **43**(6): p. 1367-78.

21. Rachid, S., et al., *Effect of Subinhibitory Antibiotic Concentrations on Polysaccharide Intercellular Adhesin Expression in Biofilm-Forming Staphylococcus epidermidis.* Antimicrobial Agents and Chemotherapy, 2000. **44**(12): p. 3357-63.

22. Conlon, K.M., H. Humphreys, and J.P. O'Gara, *icaR Encodes a Transcriptional Repressor Involved in Environmental Regulation of ica Operon Expression and Biofilm Formation in Staphylococcus epidermidis.* Journal of Bacteriology, 2002. **184**(16): p. 4400-8.

23. Jeng, W.Y., et al., *Crystal Structure of IcaR, A Repressor of the TetR Family Implicated in Biofilm Formation in Staphylococcus epidermidis.* Nucleic Acids Research, 2008. **36**(5): p. 1567-77.

24. Handke, L.D., et al., *SigmaB and SarA Independently Regulate Polysaccharide Intercellular Adhesin Production in Staphylococcus epidermidis.* Canadian Journal of Microbiology, 2007. **53**(1): p. 82-91.

25. Tormo, M.A., et al., *SarA is an Essential Positive Regulator of Staphylococcus epidermidis Biofilm Development.* Journal of Bacteriology, 2005. **187**(7): p. 2348-56.

26. Knobloch, J.K.M., et al., *Biofilm Formation by Staphylococcus epidermidis Depends on Functional RsbU, an Activator of the sigB Operon: Differential Activation Mechanisms Due to Ethanol and Salt Stress.* Journal of Bacteriology, 2001. **183**(8): p. 2624-2633.

27. Knobloch, J.K.M., et al., *RsbU-Dependent Regulation of Staphylococcus epidermidis Biofilm Formation is Mediated via the Alternative Sigma Factor Sigma(B) by Repression of the Negative Regulator Gene icaR.* Infection and Immunity, 2004. **72**(7): p. 3838-3848.

28. Rachid, S., et al., *Alternative Transcription Factor Sigma(B) is Involved in Regulation of Biofilm Expression in a Staphylococcus aureus Mucosal Isolate.* Journal of Bacteriology, 2000. **182**(23): p. 6824-6.
29. Xu, L., et al., *Role of the luxS Quorum-Sensing System in Biofilm Formation and Virulence of Staphylococcus epidermidis.* Infection and Immunity, 2006. **74**(1): p. 488-96.
30. Bischoff, M., et al., *Microarray-Based Analysis of the Staphylococcus aureus sigmaB Regulon.* Journal of Bacteriology, 2004. **186**(13): p. 4085-99.
31. Wu, S., H. de Lencastre, and A. Tomasz, *Sigma-B, A Putative Operon Encoding Alternate Sigma Factor of Staphylococcus aureus RNA Polymerase: Molecular Cloning and DNA Sequencing.* Journal of Bacteriology, 1996. **178**(20): p. 6036-42.
32. Conlon, K.M., H. Humphreys, and J.P. O'Gara, *Inactivations of rsbU and sarA by IS256 Represent Novel Mechanisms of Biofilm Phenotypic Variation in Staphylococcus epidermidis.* Journal of Bacteriology, 2004. **186**(18): p. 6208-19.
33. Vuong, C., et al., *Staphylococcus epidermidis Polysaccharide Intercellular Adhesin Production Significantly Increases During Tricarboxylic Acid Cycle Stress.* Journal of Bacteriology, 2005. **187**(9): p. 2967-73.
34. Lauble, H., et al., *The Reaction of Fluorocitrate with Aconitase and the Crystal Structure of the Enzyme-Inhibitor Complex.* Proceedings of the National Academy of Sciences of the United States of America, 1996. **93**(24): p. 13699-703.
35. Mccombie, H. and B.C. Saunders, *Fluoroacetates and Allied Compounds.* Nature, 1946. **158**(4011): p. 382-385.
36. Mack, D., N. Siemssen, and R. Laufs, *Parallel Induction by Glucose of Adherence and a Polysaccharide Antigen Specific for Plastic-Adherent Staphylococcus epidermidis: Evidence for Functional Relation to Intercellular Adhesion.* Infection and Immunity, 1992. **60**(5): p. 2048-57.
37. Sambrook, J., Fritsch, E. F., and Maniatis, T., *Molecular Cloning: A Laboratory Manual, 2nd ed.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY., 1989.
38. Novick, R.P., *Genetic Systems in Staphylococci.* Methods in Enzymology, 1991. **204**: p. 587-636.
39. Dean, B.A., et al., *Phage Typing of Coagulase-Negative Staphylococci and Micrococci.* The Journal of Hygiene, 1973. **71**(2): p. 261-70.

40. Bruckner, R., *Gene Replacement in Staphylococcus carnosus and Staphylococcus xylosus.* FEMS Microbiology Letters, 1997. **151**(1): p. 1-8.

41. Projan, S.J. and G.L. Archer, *Mobilization of the Relaxable Staphylococcus aureus Plasmid pC221 by the Conjugative Plasmid pGO1 Involves Three pC221 loci.* Journal of Bacteriology, 1989. **171**(4): p. 1841-5.

42. Greene, C., et al., *Adhesion Properties of Mutants of Staphylococcus-Aureus Defective in Fibronectin-Binding Proteins and Studies on the Expression of Fnb Genes.* Molecular Microbiology, 1995. **17**(6): p. 1143-1152.

43. Yanisch-Perron, C., J. Vieira, and J. Messing, *Improved M13 Phage Cloning Vectors and Host Strains: Nucleotide Sequences of the M13mp18 and pUC19 vectors.* Gene, 1985. **33**(1): p. 103-19.

44. Schenk, S. and R.A. Laddaga, *Improved Method for Electroporation of Staphylococcus aureus.* FEMS Microbiology Letters, 1992. **73**(1-2): p. 133-8.

45. Foster, T.J., *Molecular Genetic Analysis of Staphylococcal Virulence.* Methods in Microbiology, Vol 27, 1998. **27**: p. 433-454.

46. Southern, E.M., *Detection of Specific Sequences Among DNA Fragments Separated by Gel Electrophoresis.* Journal of Molecular Biology, 1975. **98**(3): p. 503-17.

47. Somerville, G.A., et al., *Staphylococcus aureus Aconitase Inactivation Unexpectedly Inhibits Post-Exponential-Phase Growth and Enhances Stationary-Phase Survival.* Infection and Immunity, 2002. **70**(11): p. 6373-82.

48. Majerczyk, C.D., et al., *Staphylococcus aureus CodY Negatively Regulates Virulence Gene Expression.* Journal of Bacteriology, 2008. **190**(7): p. 2257-65.

49. Gracia, E., et al., *In vitro Development of Staphylococcus aureus Biofilms Using Slime-Producing Variants and ATP-Bioluminescence for Automated Bacterial Quantification.* Luminescence, 1999. **14**(1): p. 23-31.

50. Somerville, G.A., et al., *In vitro Serial Passage of Staphylococcus aureus: Changes in Physiology, Virulence Factor Production, and agr Nucleotide Sequence.* Journal of Bacteriology, 2002. **184**(5): p. 1430-7.

51. Baughn, A.D. and M.H. Malamy, *A Mitochondrial-Like Aconitase in the Bacterium Bacteroides Fragilis: Implications for the Evolution of the*

Mitochondrial Krebs Cycle. Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(7): p. 4662-7.

52. Lowry, O.H., et al., *Protein Measurement with the Folin Phenol Reagent.* The Journal of Biological Chemistry, 1951. **193**(1): p. 265-75.

53. Delaglio, F., et al., *NMRPipe: A Multidimensional Spectral Processing System Based on UNIX Pipes.* Journal of Biomolecular NMR, 1995. **6**(3): p. 277-93.

54. Garrett, D.S., et al., *A Common-Sense Approach to Peak Picking in 2-Dimensional, 3-Dimensional, and 4-Dimensional Spectra Using Automatic Computer-Analysis of Contour Diagrams.* Journal of Magnetic Resonance, 1991. **95**(1): p. 214-220.

55. Cui, Q., et al., *Metabolite Identification via the Madison Metabolomics Consortium Database.* Nature Biotechnology, 2008. **26**(2): p. 162-4.

56. Kanehisa, M., et al., *KEGG for Linking Genomes to Life and the Environment.* Nucleic Acids Research, 2008. **36**(Database issue): p. D480-4.

57. Karp, P.D., et al., *Expansion of the BioCyc Collection of Pathway/Genome Databases to 160 genomes.* Nucleic Acids Research, 2005. **33**(19): p. 6083-9.

58. Craig, J.E., et al., *A Null Mutation in the Bacillus subtilis Aconitase Gene Causes a Block in Spo0A-Phosphate-Dependent Gene Expression.* Journal of Bacteriology, 1997. **179**(23): p. 7351-9.

59. Conlon, K.M., H. Humphreys, and J.P. O'Gara, *Regulation of icaR Gene Expression in Staphylococcus epidermidis.* FEMS microbiology letters, 2002. **216**(2): p. 171-7.

60. Collins, F.M. and J. Lascelles, *The Effect of Growth Conditions on Oxidative and Dehydrogenase Activity in Staphylococcus aureus.* Journal of General Microbiology, 1962. **29**: p. 531-5.

61. Strasters, K.C. and K.C. Winkler, *Carbohydrate Metabolism of Staphylococcus Aureus.* Journal of General Microbiology, 1963. **33**: p. 213-29.

62. Brekasis, D. and M.S. Paget, *A Novel Sensor of NADH/NAD+ Redox Poise in Streptomyces Coelicolor A3(2).* The EMBO Journal, 2003. **22**(18): p. 4856-65.

63. Lamb, H.K., et al., *The Negative Transcriptional Regulator NmrA Discriminates Between Oxidized and Reduced Dinucleotides.* The Journal of Biological Chemistry, 2003. **278**(34): p. 32107-14.

64. Gaudu, P., et al., *CcpA Regulation of Aerobic and Respiration Growth in Lactococcus lactis.* Molecular Microbiology, 2003. **50**(1): p. 183-192.

65. Jourlin-Castelli, C., et al., *CcpC, A Novel Regulator of the LysR Family Required for Glucose Repression of the citB Gene in Bacillus subtilis.* Journal of Molecular Biology, 2000. **295**(4): p. 865-78.

66. Kim, H.J., A. Roux, and A.L. Sonenshein, *Direct and Indirect Roles of CcpA in Regulation of Bacillus subtilis Krebs Cycle Genes.* Molecular Microbiology, 2002. **45**(1): p. 179-190.

67. Tramonti, A., et al., *Mechanisms of Transcription Activation Exerted by GadX and GadW at the gadA and gadBC Gene Promoters of the Glutamate-Based Acid Resistance System in Escherichia coli.* Journal of Bacteriology, 2006. **188**(23): p. 8118-27.

68. Shivers, R.P. and A.L. Sonenshein, *Activation of the Bacillus subtilis Global Regulator CodY by Direct Interaction with Branched-Chain Amino Acids.* Molecular Microbiology, 2004. **53**(2): p. 599-611.

69. Stephenson, K. and J.A. Hoch, *PAS-A Domain of Phosphorelay Sensor Kinase A: A Catalytic ATP-Binding Domain Involved in the Initiation of Development in Bacillus subtilis.* Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(26): p. 15251-6.

70. Chatterjee, I., et al., *Very Low Ethanol Concentrations Affect the Viability and Growth Recovery in Post-Stationary-Phase Staphylococcus aureus Populations.* Applied and Environmental Microbiology, 2006. **72**(4): p. 2627-2636.

71. Kornmann, H., et al., *Influence of Residual Ethanol Concentration on the Growth of Gluconacetobacter xylinus I 2281.* Applied Microbiology and Biotechnology, 2003. **62**(2-3): p. 168-173.

72. Guedon, E., et al., *Pleiotropic transcriptional Repressor CodY Senses the Intracellular Pool of Branched-Chain Amino Acids in Lactococcus lactis.* Molecular Microbiology, 2001. **40**(5): p. 1227-1239.

73. Sonenshein, A.L., *CodY, A Global Regulator of Stationary Phase and Virulence in Gram-Positive Bacteria.* Current Opinion in Microbiology, 2005. **8**(2): p. 203-7.

74. Ratnayake-Lecamwasam, M., et al., *Bacillus subtilis CodY Represses Early-Stationary-Phase Genes by Sensing GTP Levels.* Genes & Development, 2001. **15**(9): p. 1093-103.

75. Henkin, T.M., et al., *Catabolite Repression of Alpha-Amylase Gene Expression in Bacillus subtilis Involves a Trans-Acting Gene Product Homologous to the Escherichia coli lacl and galR Repressors.* Molecular Microbiology, 1991. **5**(3): p. 575-84.

76. Seidl, K., et al., *Staphylococcus aureus CcpA Affects Biofilm Formation.* Infection and Immunity, 2008. **76**(5): p. 2044-50.

77. Sonenshein, A.L., *Control of Key Metabolic Intersections in Bacillus subtilis.* Nature Reviews. Microbiology, 2007. **5**(12): p. 917-27.

78. Somerville, G., C.A. Mikoryak, and L. Reitzer, *Physiological Characterization of Pseudomonas aeruginosa During Exotoxin A Synthesis: Glutamate, Iron Limitation, and Aconitase Activity.* Journal of Bacteriology, 1999. **181**(4): p. 1072-8.

79. Somerville, G.A., et al., *Correlation of Acetate Catabolism and Growth Yield in Staphylococcus aureus: Implications for Host-Pathogen Interactions.* Infection and Immunity, 2003. **71**(8): p. 4724-32.

80. Throup, J.P., et al., *The srhSR Gene Pair from Staphylococcus aureus: Genomic and Proteomic Approaches to the Identification and Characterization of Gene Function.* Biochemistry, 2001. **40**(34): p. 10392-401.

81. Varghese, S., Y. Tang, and J.A. Imlay, *Contrasting Sensitivities of Escherichia coli Aconitases A and B to Oxidation and Iron Depletion.* Journal of Bacteriology, 2003. **185**(1): p. 221-30.

82. Kohanski, M.A., et al., *A Common Mechanism of Cellular Death Induced by Bactericidal Antibiotics.* Cell, 2007. **130**(5): p. 797-810.

83. Tomlins, R.I., M.D. Pierson, and Z.J. Ordal, *Effect of Thermal Injury on the TCA Cycle Enzymes of Staphylococcus aureus MF 31 and Salmonella Typhimurium 7136.* Canadian Journal of Microbiology, 1971. **17**(6): p. 759-65.

CHAPTER 6

USING NMR METABOLOMICS TO INVESTIGATE TRICARBOXYLIC ACID CYCLE-DEPENDENT SIGNAL TRANSDUCTION IN *STAPHYLOCOCCUS EPIDERMIDIS*[§]

6.1 Introduction

Staphylococcus epidermidis is a skin-resident, opportunistic pathogen that is the leading cause of hospital-associated infections [1]. Although the type and severity of diseases produced by this bacterium varies, its most common infectious manifestation is associated with implanted biomaterials. The dramatic environmental changes that occur during the transition from being skin-resident to residing on implanted biomaterials necessitates the need for changes in the expression of genes coding for enzymes required for growth in the new environment. This environmental adaptation often includes activating transcription of virulence genes; hence, most virulence genes are regulated by environmental and nutritional signals [2]. Accordingly, a major area of interest in microbiology is determining how bacteria "sense" and respond to environmental signals. Given the tremendous diversity of microbial life, it is not surprising that the mechanisms bacteria employ are equally diverse. These mechanisms include two-component regulatory systems, alternative σ factors, mechanosensors, small RNAs, riboswitches, and

[§] Chapter 6 was adapted from Sadykov, M., et. al., Using NMR Metabolomics to Investigate Tricarboxylic Acid Cycle-Dependent Signal Transduction in *Staphylococcus epidermidis*, *Journal of Biological Chemistry* (2010), **285**, 36616-36624. Reprinted with permission, copyright 2010 American Society for Biochemistry and Molecular Biology.

many others. Although remarkable advances have been made in identifying the response regulators, our knowledge of signaling mechanisms has lagged behind, the exception being cell-density signaling.

The tricarboxylic acid (TCA) cycle has been implicated as regulating or affecting staphylococcal virulence and/or virulence determinant biosynthesis [3-9]. The TCA cycle has three primary functions: (i) to provide biosynthetic intermediates, (ii) to generate reducing potential, and (iii) to directly produce a small amount of ATP. The availability of biosynthetic intermediates affects the availability of amino acids and nucleic acids. Increasing the reducing potential alters the bacterial redox balance, necessitating oxidation reactions via the electron transfer chain. The small amount of ATP produced directly by the TCA cycle is amplified many times when the ATP generated by oxidative phosphorylation is considered. In short, the TCA cycle has a central function in maintaining the bacterial metabolic status. Importantly, the activity of TCA cycle enzymes is affected by the availability of nutrients and a variety of stress inducing stimuli [9-12]; thus, the availability of biosynthetic intermediates, the redox status, and the energy status can be altered by nutritional and environmental stimuli. These observations led us to propose a fourth function for the TCA cycle, the transduction of external signals into intracellular metabolic signals that can be “sensed” by metabolite-responsive regulatory proteins [2]. Fundamental to this hypothesis are the predictions that disparate environmental stimuli will cause common metabolic changes and that these metabolic changes will precede regulatory changes.

Two of the more extensively studied environmental stimuli that influence *S. epidermidis* virulence determinant biosynthesis are iron limitation [13, 14] and ethanol stress [15, 16]. The effect of iron limitation on bacterial growth is primarily through preventing the activity of enzymes that require iron as a cofactor and altering transcription of iron-regulated genes [2]. As *S. epidermidis* has many iron-requiring enzymes (*e.g.* aconitase, serine dehydratase, peptide deformylase, iron-containing alcohol dehydrogenase, nitrate reductase, etc.), it is reasonable to expect that the metabolic effects of iron-limited growth are diverse and not restricted to the TCA cycle. Ethanol denatures proteins in the cytoplasmic membrane, causing changes in membrane permeability, which can lead to the loss of membrane integrity [17]. With the exception of the succinate dehydrogenase complex, most TCA cycle enzymes are not membrane-associated; hence, it is reasonable to predict that the deleterious effects of ethanol stress are largely independent of the TCA cycle. Taken together, these observations suggest that disparate environmental conditions will cause divergent metabolomic changes. In contrast to this suggestion, our central hypothesis predicts that different stresses will cause common metabolomic changes that are dependent on the TCA cycle. To test our central hypothesis, we chose to induce environmental stress by growing bacteria in an iron-limited medium or in a medium containing ethanol and assessing the metabolic changes using NMR metabolomics.

6.2 Experimental Procedures

6.2.1 Bacterial Strains, Media, and Growth Conditions

S. epidermidis wild-type strain 1457 [18] and the isogenic aconitase mutant strain 1457-*acnA*::*tetM* (*tetM* cassette inserted into position 856 of the 2,702-bp *acnA* gene) and σ^B mutant strain 1457-*sigB*::*dhfr* [7, 19] have been described. Strains 1457-*codY*, 1457-*ccpA*, 1457-*acnA*-*codY*, and 1457-*acnA*-*ccpA* were constructed using the gene splicing by overlap extension (gene SOEing) technique [20] to replace the gene of interest with an antibiotic resistance marker (*i.e. ermB* or *tetM*). Primers were designed to amplify ~1-kb regions upstream and downstream of the gene of interest based on the genome sequence of *S. epidermidis* strain RP62A. Gene knockouts were confirmed by PCR and Southern blot hybridization. In addition, strains containing mutations in the single *S. epidermidis* aconitase gene were assayed to ensure that no aconitase activity was detected (data not shown). All strains were grown in tryptic soy broth without dextrose (TSB;4 BD Biosciences) supplemented with 0.25% glucose (Sigma) or 0.25% $^{13}C_6$-glucose (Cambridge Isotope Laboratories). Deferrated TSB (DTSB) was prepared by adding 50 g of Chelex 100 (Sigma) to ~1 liter of TSB and stirring at 4 °C for 20 h. After 20 h, the Chelex resin was removed, 1mM $MgSO_4$ was added, the volume was adjusted to 1 liter, and the medium was filter-sterilized. To induce ethanol stress and minimize growth defects, ethanol or deuterated ethanol (Isotec) was added to the medium at a final concentration of 4% (v/v). All cultures were inoculated 1:200 from overnight cultures (normalized for growth) into glucose supplemented TSB, incubated at 37 °C, and aerated at 225 rpm with a flask-to-medium ratio of 7:1. Bacterial growth was assessed by

measuring the optical density at 600 nm (A_{600}). Antibiotics, when used, were purchased from Fisher Scientific or Sigma and used at the following concentrations: chloramphenicol (8 μg/mL), trimethoprim (10 μg/mL), and erythromycin (8 μg/mL).

6.2.2 Aconitase Activity Assay

Cell-free lysates of *S. epidermidis* were prepared as follows. Aliquots (3 mL) were harvested by centrifugation (1 min at 20,800 x *g*) at the indicated times, suspended in 1.5 mL of lysis buffer containing 90 mM Tris (pH 8.0) and 100 μM fluorocitrate. The samples were lysed in 2-mL screw cap tubes containing lysing matrix B using a FastPrep instrument (MP Biomedicals). The lysate was centrifuged for 5 min at 20,800 x *g* at 4 °C. Aconitase activity in the resulting cell-free lysate was assayed by the method described by Kennedy *et al.* [21]. One unit of aconitase activity is defined as the amount of enzyme necessary to give a $\Delta A240$ min^{-1} of 0.0033 [22]. Protein concentrations were determined by the Lowry method [23].

6.2.3 Northern Blot Analysis

Northern blot analysis of transcripts was performed as described [7]. Oligonucleotide primers used in making DNA probes were designed using the *S. epidermidis* RP62A genome sequence. Probes for Northern blotting were generated by PCR amplification of unique internal regions of RNAIII and *glnA* (*femC*) genes using the following primers: femC, forward, 5′-

GATGTTTGATGGTTCATCTATTGAAGGTTTCG-3′; femC, reverse, 5′-GCAGTATCAGTCAATTGTAAATCACCTTCAG-3′; RNAIII, forward, 5′-TGAAAAATTTGCTTAATCTAGTCGAGTG-3′; and RNAIII, reverse, 5′-CATGATAAATTGAATGTTGTTTACGATAGC-3′. DNA probes were labeled using the North2South random prime labeling kit (Pierce). Electrophoresis, transfer of the RNA to the Nytran SPC nylon membrane (Whatman), and hybridization were done using the NorthernMax kit (Ambion). Detection was performed using the chemiluminescent nucleic acid detection module (Pierce).

6.2.4 PIA Immunoblot Assay

PIA accumulation was determined as described [24].

6.2.5 NMR Sample Preparation

NMR samples for one-dimensional ^{1}H spectra were prepared from 10 independent, 25mL *S. epidermidis* cultures. Two-dimensional ^{1}H-^{13}C HSQC [25, 26] and two-dimensional ^{1}H-^{1}H TOCSY [27] spectra were prepared from three independent 50mL cultures. The TSB medium used in the two-dimensional ^{1}H-^{13}C HSQC analysis contained 0.25% $^{13}C_6$-glucose (Cambridge Isotope Laboratories). For two-dimensional ^{1}H-^{13}C HSQC and two-dimensional ^{1}H-^{1}H TOCSY involving ethanol stress, deuterated ethanol (Isotec) was used to minimize the contribution of exogenous ethanol to the NMR spectra. For one-dimensional ^{1}H NMR experiments, 2.74 A_{600} units were harvested at each time point, and for the two-dimensional ^{1}H-^{13}C HSQC and two-dimensional ^{1}H-^{1}H

TOCSY experiments, 5.48 A_{600} units were collected. Bacteria were harvested by centrifugation (4,000 rpm for 5 min), suspended in 50 mM phosphate buffer in 100% D_2O at pH 7.2 (uncorrected), and lysed using lysing matrix B tubes and a FastPrep instrument. The lysates were centrifuged to remove cell debris and glass beads and then frozen in liquid nitrogen. All samples were kept at -80 °C until ready for analysis. At the time of use, a 600 µL aliquot of the cell-free lysate was transferred to each NMR tube.

6.2.6 NMR Analysis

The NMR spectra were collected on a Bruker 500 MHz Avance spectrometer equipped with a triple resonance, *z* axis gradient cryoprobe. A BACS-120 sample changer with Bruker Icon software was used to automate the NMR data collection. The one-dimensional ^{1}H NMR spectra collection and principal component analysis (PCA) were performed as described with minor modifications [28-30]. Briefly, each multidimensional NMR spectrum (chemical shifts and peak intensities) was converted to a single point in a multidimensional Cartesian space. Conceptually, each axis corresponds to a specific chemical shift, where the peak intensity is the value along the axis. PCA identifies a principal component vector (P_1) corresponding to the largest variation in the data set within this multidimensional space. The second vector (P_2) is orthogonal to the first and represents the next largest variation in the data set. Each successive vector describes a diminishing amount of the variability of the data set, where most of the variability is described by the first two principal components. The PC1 and PC2 scores (unitless values) are effectively the individual fit of each NMR spectrum to P_1 and P_2. The PC1

and PC2 scores are usually presented in a two-dimensional plot, where similar NMR spectra cluster together.

Solvent presaturation used excitation sculpting to efficiently remove the solvent and maintain a flat baseline, eliminating any need for baseline collection that may induce artifacts in the two-dimensional scores plot [31]. Each NMR spectrum was center-averaged for PCA to minimize any experimental variations between cultures [32].

Two-dimensional ^{1}H-^{13}C HSQC spectra were collected and processed as described previously [7]. Two-dimensional ^{1}H-^{1}H TOCSY spectra were collected with WATERGATE solvent presaturation [33] and a relaxation delay of 2 s. A total of 1,024 data points with a sweep width of 5,000 Hz and 256 data points with a sweep width of 5,001.324 Hz were collected in the direct and indirect ^{1}H dimensions, respectively. A total of 16 dummy scans and 8 acquisition scans were used to obtain each of the two-dimensional ^{1}H-^{1}H TOCSY NMR spectra. The two-dimensional ^{1}H-^{1}H TOCSY NMR spectra were processed similar to the two-dimensional ^{1}H-^{13}C HSQC spectra, and both spectra were analyzed using NMRView (One Moon Scientific [34]) and Sparky [35] to identify chemical shifts and assign peak intensities.

The observed NMR peaks in the two-dimensional ^{1}H-^{13}C HSQC and ^{1}H-^{1}H TOCSY spectra were assigned to specific metabolites using ^{1}H and ^{13}C chemical shift tolerances of 0.05 and 0.50 ppm, respectively, and the Madison Metabolomics Consortium Database (MMCD) [36], the BioMagResBank [37], and the Human Metabolome Database [38]. The presence of metabolites and metabolic pathways were

verified with the Kyoto Encyclopedia of Genes and Genomes (KEGG) [39] and the MetaCyc [40] databases.

Peak intensities were normalized for each two-dimensional NMR spectrum by dividing by the average peak intensity. The triplicate data sets were then used to calculate average intensities for each peak observed in the two-dimensional spectra for strain 1457, 1457-*acnA*, ethanol stress, and iron limitation. A percentage of error was calculated for each peak by dividing the standard deviation by the average peak intensity. The average peak intensities were then used to calculate a percentage of difference relative to the wild-type bacteria in TSB medium. Peaks with calculated percentage of differences greater than five times the average percentage of error were considered to have either decreased or increased concentrations relative to the wild-type strain 1457. Peaks with less than a 5-fold deviation were considered similar. Secondary peaks assigned to the same metabolite were required to have the same relative change in intensity to be classified as a metabolite with an increase or decrease in concentration.

6.2.7 Metabolomic Dendrogram

The relative clustering patterns in the PCA two-dimensional scores plots were quantitatively analyzed using a tree diagram and bootstrapping technique [41]. The PC1 and PC2 scores for each set of 10 duplicate NMR spectra representing a specific metabolic state (iron limitation, ethanol treatment, etc.) were used to calculate an average PC score and standard deviation. Any PC scores outside 2 standard deviations were removed, and a new average was calculated. The average PC scores represent the center

of a cluster of NMR spectra (metabolic state) in the two-dimensional scores plot. The process is repeated for each set of 10 duplicate NMR spectra. Distances between the average PC positions for each metabolic state are then calculated using the standard equation for a Euclidean distance to create a distance matrix.

To assess the significance of the similarity (overlap) or difference (separation) observed between pairs of clusters in the two dimensional scores plot, standard bootstrapping methods were also applied [42, 43]. Briefly, the average PC scores were recalculated by randomly selecting points from the data set. Distances were recalculated between the clusters using the new average PC scores to create a new distance matrix. The process was repeated until 100 different distance matrices were created and transferred to version 3.68 of the PHYLIP [44] suite of software programs. PHYLIP calculates a tree for each distance matrix and then determines a consensus tree. The program calculates a bootstrap value for each node, which is simply the number of times the node appears in all 100 trees. Bootstrap values below 50% imply a statistically insignificant separation. Conversely, as the bootstrap number increases above 50%, the confidence in the tree branch or separation increases.

6.3 Results

6.3.1 Disparate Environmental Stresses Create a Metabolic Block in the TCA Cycle

To determine whether ethanol stress and iron limitation alter TCA cycle activity, the specific activity of aconitase in *S. epidermidis* strain 1457 at 2 (exponential growth) and 6 h (post-exponential growth) after inoculation was assessed (Figure 6.1). As

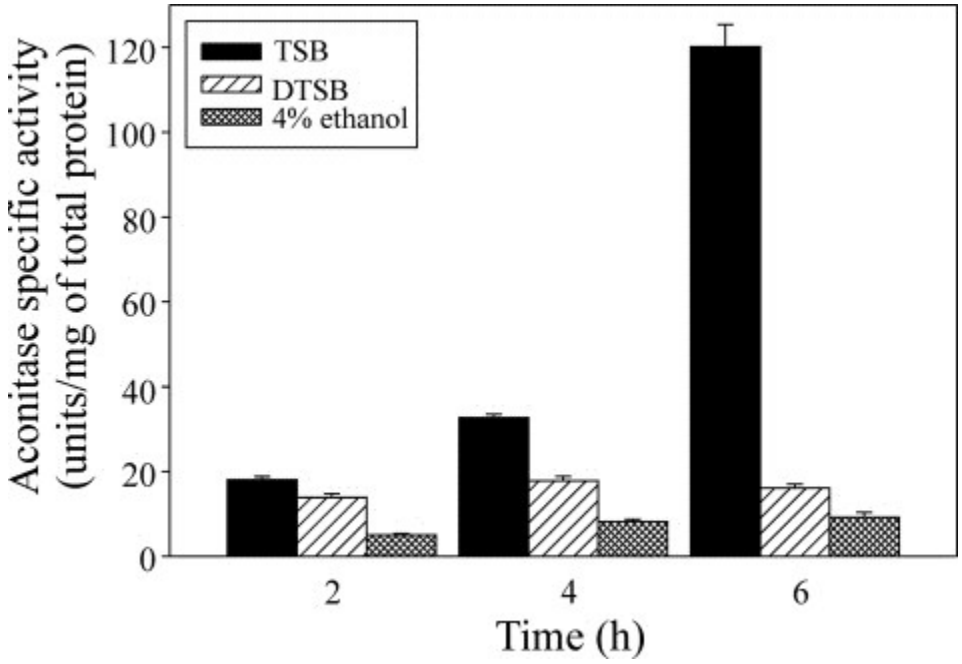


Figure 6.1. The temporal induction of aconitase-specific activity is inhibited by dissimilar stressors. Aconitase activity was assessed during the exponential (2 and 4 h) and post-exponential (6 h) growth phases during growth in DTSB or TSB containing 4% ethanol. The data are presented as the mean and S.E. of two independent experiments each determined in triplicate.

expected, iron-limited growth and ethanol stress prevented the post-exponential growth phase increase in the specific activity of the iron-requiring enzyme aconitase, creating a metabolic block in the TCA cycle (Figure 6.1). Although ethanol-stressed bacteria are in the post-exponential growth phase at 6 h after inoculation, their growth is slower, which slows the consumption of glucose, and excess glucose can repress transcription of TCA cycle genes. Irrespective of the mechanism by which ethanol repressed aconitase-specific activity, the normal post-exponential growth phase increase in TCA cycle activity did not occur. These data demonstrate that environmental stresses whose deleterious effects are substantially different from one another have a similar effect on TCA cycle function.

6.3.2 Environmental Stimuli Elicit TCA Cycle-dependent Metabolic Changes

The TCA cycle provides biosynthetic intermediates, ATP, and reducing potential; therefore, alteration of TCA cycle activity will alter the metabolic status of a bacterium. To determine the metabolic changes associated with iron limitation, ethanol stress, and TCA cycle inactivation, NMR metabolomic analysis (28, 29) was used to assess the stressed and non-stressed metabolomes of strains 1457 and the TCA cycle inactive strain 1457-*acnA*. Specifically, *S. epidermidis* strains 1457 and 1457-*acnA* were grown for 2 or 6 h in TSB, TSB with 4% ethanol, or DTSB. Following acquisition of the NMR spectra, the table of integrals was used for PCA (Figure 6.2a). As expected, during the exponential growth phase, PCA revealed that the effects of ethanol stress and iron limitation on the metabolome were largely independent of the TCA cycle (Figure 6.3). This was expected due to the normal repression of TCA cycle activity during nutrient-

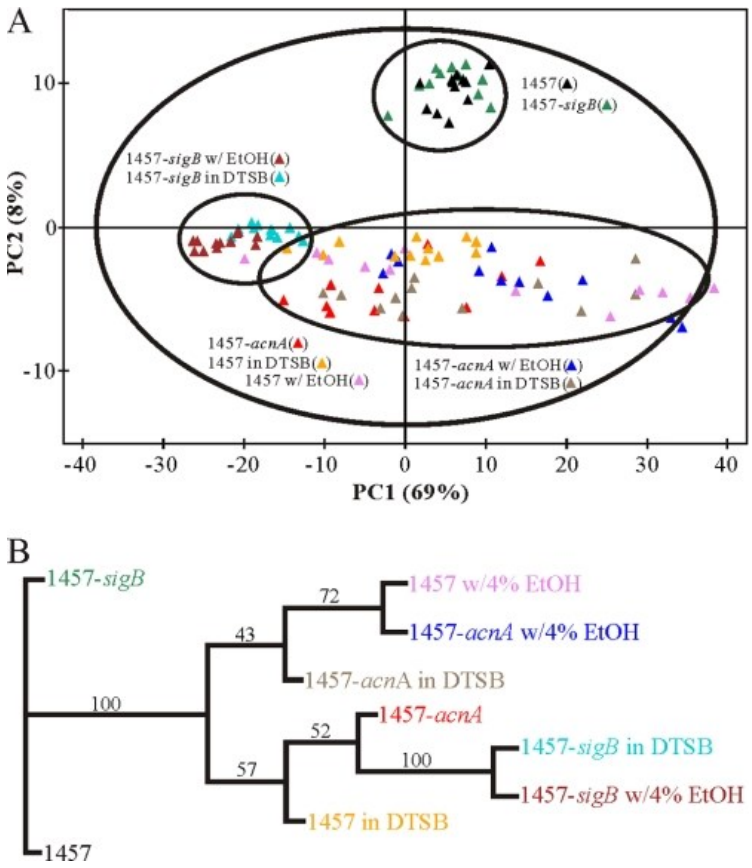


Figure 6.2 Environmental stressors cause metabolomic perturbations similar to TCA cycle inactivation. **a)** PCA two-dimensional scores plot comparing non-stressed, ethanol-stressed, or iron-limited cultures of strains 1457, 1457-*acnA*, and 1457-*sigB*::*dhfr* grown for 6 h. Symbols and colors are defined in the figure. The *ovals* are manually drawn to identify clusters of related samples and to guide the reader. They are not statistically relevant. The relative contribution of each principal component is indicated in the *parentheses*. **b)** metabolic tree generated using the PCA scores plot data demonstrating the relationship between stresses and strains. Bootstrap values are indicated on the dendrogram branches. Bootstrap values below 50% imply a statistically insignificant separation; conversely, as the bootstrap number increases above 50%, the confidence in the tree branch or separation increases.

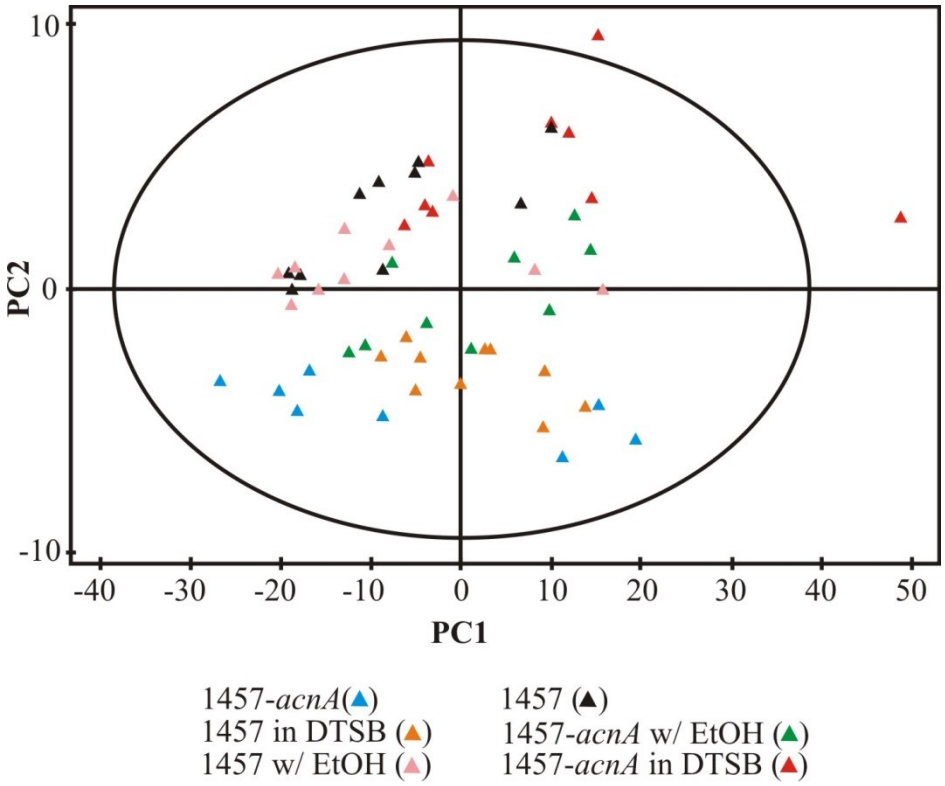


Figure 6.3. PCA 2D scores plot comparing non-stressed, ethanol stressed, or iron-limited cultures of strains 1457, 1457-*acnA*, and 1457-*sigB::dhfr* grown for 2h. Symbols and colors are defined in the figure.

Table 6.1. Metabolites that have increased or decreased concentrations relative to the wild-type strain 1457 at **2 h** after inoculation.

Strain 1457[a] (DTSB) **Iron-limitation**	Strain 1457[a] (4% Ethanol) **Ethanol stress**	Strain 1457-acnA[a] (TSB) **TCA cycle inactivation**
Metabolites whose concentration is **increased** relative to strain 1457 grown in TSB medium.		
Acetyl-phosphate	Acetyl-phosphate	Acetaldehyde Acetyl-phosphate Citrate Glyceraldehyde
Metabolites whose concentration is **decreased** relative to strain 1457 grown in TSB medium.		
Glutamine Succinate	Lactate Glutamine Succinate	Acetyl-glutamate[b] Asparagine Glutamate Lactate Glutamine Succinate

[a] The intracellular concentration was considered to be increased or decreased when the percentage of difference in the NMR peak intensities was 5-fold greater than the percentage of error observed in the peak intensities between triplicate NMR spectra.

[b] Due to peak overlap, we are unable to determine whether the metabolite is acetylglutamine or acetyl-glutamate; however, we note that acetyl-glutamine is uncommon in prokaryotes.

Table 6.2. Metabolites that have increased or decreased concentrations relative to the wild-type strain 1457 at **6 h** after inoculation.

Strain 1457[a] (DTSB) **Iron-limitation**	Strain 1457[a] (4% Ethanol) **Ethanol stress**	Strain 1457-acnA[a] (TSB) **TCA cycle inactivation**
Metabolites whose concentration is **increased** relative to strain 1457 grown in TSB medium		
Acetaldehyde	Acetaldehyde	Acetaldehyde
Acetate	Acetate	Acetate
Acetylalanine		
N-Acetyl-glucosamine	N-Acetyl-glucosamine	
	Aceytl-glucosamine 6-phosphate	
N-Acetyl-mannosamine	N-Acetyl-mannosamine	N-Acetyl-mannosamine
N-Acetyl-neuraminate	N-Acetyl-neuraminate	N-Acetyl-neuraminate
Acetyl-phosphate	Acetyl-phosphate	Acetyl-phosphate
		Arginine
		Citrate
Ethanol		
Glucosamine	Glucosaminec	Glucosamine
		Galactose-1-phosphate
GDP	GDP	
Glucose	Glucose	Glucose
Glucose-1-phosphate		
Glucose-6-phosphate	Glucose-6-phosphate	Glucose-6-phosphate
Glyceraldehyde	Glyceraldehyde	Glyceraldehyde
Lactate	Lactate	Lactate
	Myo-inositol	
		Proline
Ribose[c]	Ribose	Ribose
	UDP-N-acetyl-glucosamine	
Metabolites whose concentration is **decreased** relative to strain 1457 grown in TSB medium		
α-ketoglutarate	α-ketoglutarate	α-ketoglutarate
γ-aminobutryate	γ-aminobutyrate	γ-aminobutyrate
Acetyl-glutamate[b]	Acetyl-glutamate[b]	Acetyl-glutamate[b]
		Acetyl-ornithine
		Alanine
Arginine		
Asparagine	Asparagine	Asparagine
		Aspartate
B-alanine	B-alanine	B-alanine
Citrulline	Citrulline	Citrulline
	Ethanol	Ethanol
Fructose-6-phosphate	Fructose-6-phosphate	Fructose-6-phosphate
Glutamate	Glutamate	Glutamate

Table 6.2. Continued

Glutamine	Glutamine	Glutamine
Isocitrate	Isocitrate	Isocitrate
Methionine	Methionine	Methionine
NAD+	NAD+	
Ornithine	Ornithine	Ornithine
O-succinyl-L-homoserine	O-succinyl-L-homoserine	O-succinyl-L-homoserine
Proline	Proline	
S-Adenosyl-L-methionine	S-Adenosyl-L-methionine	S-Adenosyl-L-methionine
Sedheptulose		
Selenomethionine	Selenomethionine	Selenomethionine
UDP-N-acetyl-glucosamine		

[a] The intracellular concentration was considered to be increased or decreased when the percentage of difference in the NMR peak intensities was 5-fold greater than the percentage of error observed in the peak intensities between triplicate NMR spectra.

[b] Due to peak overlap, we are unable to determine whether the metabolite is acetylglutamine or acetyl-glutamate; however, we note that acetyl-glutamine is uncommon in prokaryotes.

[c] The percentage of difference in the NMR peak intensities of these metabolites fell just below the 5-fold cutoff in the percentage of error observed in the peak intensities between the triplicate NMR spectra.

rich growth (Figure 6.1) [8, 45]. Despite the TCA cycle being repressed during the exponential growth phase, the different stresses induced common metabolomic changes (Table 6.1). In contrast to the exponential growth phase, PCA of post-exponential growth phase metabolomes revealed that ethanol stress and iron limitation induced metabolomic changes very similar to TCA cycle inactivation (Figure 6.2a and Table 6.2). In addition, these data highlight the relative insensitivity of the metabolome of strain 1457-*acnA* to ethanol stress and iron-limited growth, confirming that the major effect of these stressors is dependent upon the TCA cycle. That being said, the more diffuse clustering of ethanol stressed metabolomes of both the wild-type and the aconitase mutant strains suggest that ethanol stress had TCA cycle-independent metabolomic effects (Figure 6.2a). The TCA cycle-independent effects are likely due to the denaturation of membrane proteins not related to electron transport or the TCA cycle. Taken together, these data demonstrate that diverse environmental stimuli elicit common metabolic changes that require the TCA cycle.

The common metabolomic response to environmental stimuli can be more easily observed by a recently developed method to visualize PCA data [41]. By calculating an average position for each data set, such that each PC value (PC1, PC2, etc.) is treated as an axis in a Cartesian coordinate system, a distance matrix can be generated. Correspondingly, methods developed for representing genetic distances in phylogenetic trees can be used to create a metabolomic dendrogram [44] (Figure 6.2b). Using this approach, it becomes clear that stress-induced metabolomics responses are very similar to the metabolome of the aconitase-deficient strain 1457-*acnA*. As with the two-dimensional

scores plot (Figure 6.2a), the higher bootstrap values in the dendrogram for ethanol-stressed cultures also indicate that ethanol has TCA cycle-independent effects on the metabolome.

S. epidermidis grown in TSB under aerobic conditions have two distinct metabolic states: the nutrient-rich exponential phase and the nutrient-limited post-exponential phase. The transition from nutrient-rich conditions to nutrient-limited growth coincides with the transition from generating ATP by substrate-level phosphorylation to using oxidative phosphorylation. The reduced dinucleotides that drive oxidative phosphorylation are primarily derived from the TCA cycle; thus, inhibiting TCA cycle activity (Figure 6.1) hinders the transition to oxidative phosphorylation and the post-exponential growth phase [8, 45]. Iron-limited growth of strain 1457 or aconitase inactivation did not significantly alter the growth rate, although aconitase inactivation did increase the lag phase (data not shown). Both aconitase inactivation and iron-limited growth caused an early entry into the stationary phase; as such, the growth yield was decreased. As stated, the addition of 4% ethanol decreased the growth rate; therefore, it slowed the consumption of glucose. Based on these observations, it was reasonable to hypothesize that post-exponential growth phase (6 h)-stressed metabolomes will be more similar to an unstressed exponential growth phase (2 h) metabolome than to the unstressed metabolome of post-exponential growth phase of cultures. As expected, PCAs of unstressed strain 1457 cultures, grown for 2 or 6 h, form separate subsets in a three-dimensional scores plot (Figure 6.4). Consistent with our hypothesis, PCAs of post-exponential growth phase-stressed and *acnA* mutant cultures were more closely

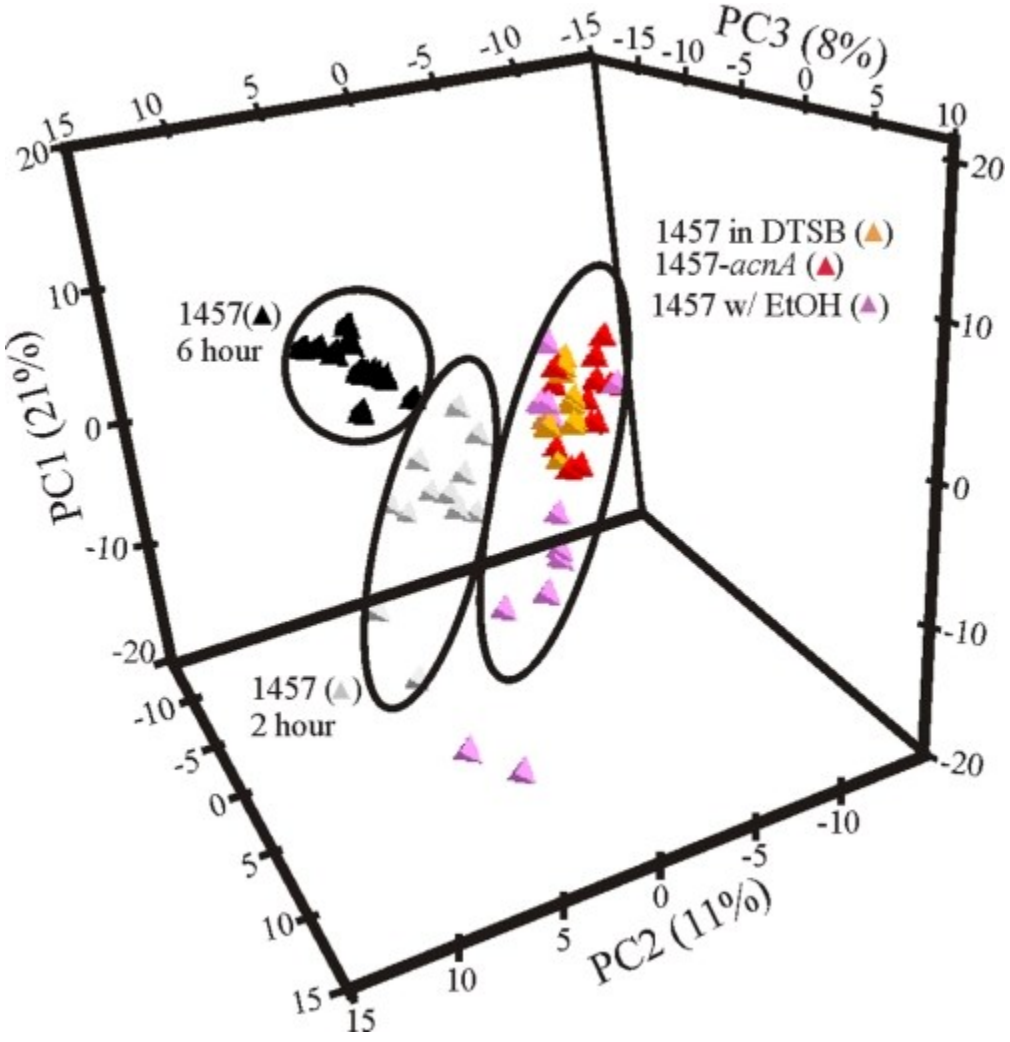


Figure 6.4. PCA three-dimensional scores plot comparing non-stressed, ethanol-stressed, or iron-limited cultures of strains 1457 grown for 6 h with that of strain 1457 grown for 2h. Symbols and colors are defined in the figure. The *ovals* are manually drawn to identify clusters of related samples and to guide the reader. They are not statistically relevant. The relative contribution of each principal component is indicated within the *parentheses*.

associated with the unstressed exponential growth phase metabolome of strain 1457 than with the strain 1457 post-exponential growth phase metabolome (Figure 6.4). These data suggest that any stress that interferes with TCA cycle function results in a metabolome similar to an unstressed exponential phase culture.

6.3.3 *Metabolomic Changes Are Largely Independent of the σ^B-mediated General Stress Response*

In staphylococci, σ^B controls the general stress response and as such is activated during stress conditions, growth phase transitions, and morphological changes [46, 47]. As stated previously, the regulation of many virulence determinants is affected by environmental stresses; therefore, the stress-dependent activation of σ^B has been an important area of research into the environmental regulation of staphylococcal virulence determinants [48-50]. Interestingly, σ^B does not directly respond to most environmental signals, suggesting another path to transduce stress signals that is independent of σ^B. To test this possibility, an *S. epidermidis sigB* mutant strain (1457-*sigB*::*dhfr*) [19] was grown for 2 or 6 h in TSB, TSB with 4% ethanol, or DTSB and the metabolomes were analyzed by NMR spectroscopy. The PCA scores plot demonstrates that the majority of metabolomics changes associated with iron limitation and ethanol stress occur largely independent of σ^B (Figure 6.2a). Additionally, the metabolomic dendrogram confirms that the stressor-induced metabolic changes observed in strain 1457-*sigB*::*dhfr* are most closely associated with those in the TCA cycle mutant strain 1457-*acnA* (Figure 6.2b). Taken together, these data demonstrate that environmental stresses can alter the

staphylococcal metabolome by a largely σ^B-independent mechanism that requires the TCA cycle.

6.3.4 *Metabolomic Changes Precede Genetic Changes*

Metabolomic data demonstrate that ethanol stress, iron limitation, and TCA cycle inactivation decrease the intracellular concentration of Gln relative to the wild-type strain grown in TSB medium (Tables 6.1 and 6.2). The two more likely explanations for the decreased intracellular concentration of Gln are: (i) the stressors alter enzymatic activity, causing a decrease in the concentration of Gln; or (ii) the stressors decrease transcription of genes involved in the biosynthesis of Gln, resulting in a decreased concentration of Gln. If the first possibility is correct, then stressors will cause an increase in the transcription of Gln biosynthetic genes as bacteria attempt to compensate for the decreased availability of Gln. If the second possibility is correct, then stressors will cause a decrease in the transcription of Gln biosynthetic genes. To determine which of these two possibilities was correct, we performed Northern blot analysis on glutamine synthetase (*femC*; also known as *glnA*) (Figure 6.5). The data suggest that the first possibility is the more correct one; specifically, bacteria are responding to metabolomics changes by increasing transcription of genes necessary to counterbalance those changes. Interestingly, in untreated wild-type cultures, the post-exponential growth phase concentration of Glu and Gln increased between two and five times that of the exponential growth phase concentration (data not shown), and this increase correlated with a post-exponential growth phase decrease in *glnA* mRNA levels (Figure 6.5).

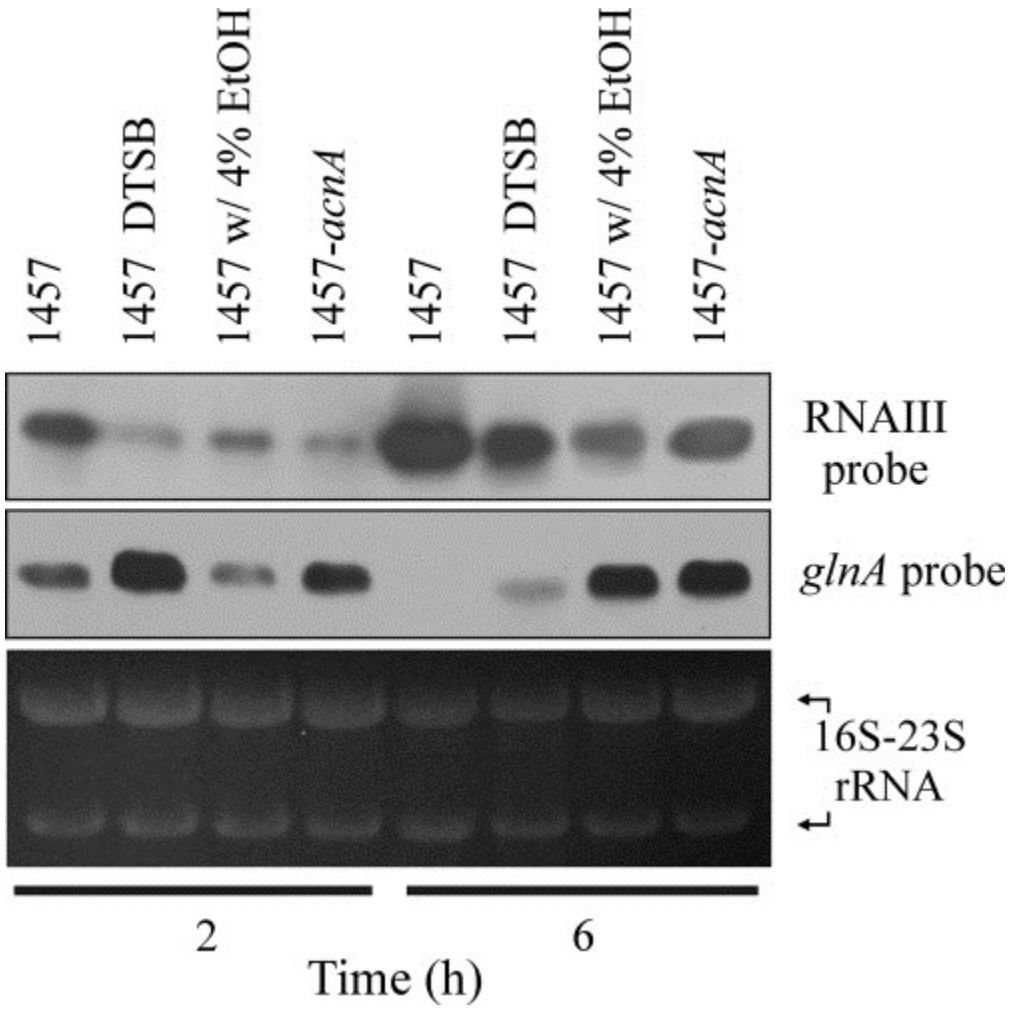


Figure 6.5. Northern blot analysis of RNAIII and *glnA* mRNA levels in the exponential (2 h) and post-exponential (6 h) phases of growth. To ensure that equivalent quantities of RNA were loaded in the gel, 23 S and 16 S rRNA were visualized by ethidium bromide staining and used as loading controls (*bottom panel*). The results are representative of at least two independent experiments.

Similarly, Gln and Glu were not detected in the NMR spectra of the aconitase mutant strain, and this correlated with a high level of *glnA* mRNA in both the exponential and the post-exponential growth phases.

The correlation between Gln and Glu concentrations and *glnA* mRNA levels is consistent with a GlnR-dependent regulation of *glnA* transcription [2]. This correlation was maintained for Gln/Glu-sufficient or -insufficient conditions; however, the intermediate concentrations of Gln and Glu found during ethanol stress and iron-limited growth (data not shown) produce mixed *glnA* mRNA levels (Figure 6.5). These data suggest that for the concentrations of Gln and Glu to affect *glnA* transcription, the stress-induced concentration change must be sufficiently large.

6.3.5 CcpA Responds to TCA Cycle-associated Metabolomic Changes

Ethanol stress, iron limitation, and TCA cycle inactivation increased the post-exponential growth phase concentration of glucose-6-phosphate (Table 6.2). Glycolytic intermediates such as glucose-6-phosphate and fructose-1,6-bisphosphate increase the ATP-dependent phosphorylation of the histidine-containing protein (HPr) by enhancing the activity of the HPr kinase [51]. The increase in phosphorylated HPr enhances its interaction with the catabolite control protein A (CcpA) [52-54]. CcpA primarily functions as a repressor; however, it also activates transcription of genes involved in fermentation and overflow metabolism [53, 55]. In addition to the concentration of glucose-6-phosphate being increased by TCA cycle stress, the concentrations of several fermentation products or intermediates (*i.e.* lactate, acetate, and acetaldehyde) and the

small phosphordonor acetyl phosphate (an indicator of overflow metabolism) were also increased (Table 6.1), consistent with a change in CcpA-mediated regulation. The repressor CodY also contributes to the regulation of overflow metabolism [55]; however, CodY responds to the intracellular concentrations of branched chain amino acids [56]. TCA cycle stress did not alter the concentrations of branched chain amino acids beyond the 5-fold threshold (Table 6.1), suggesting that the increase in overflow metabolism was independent of CodY.

In *Staphylococcus aureus*, CcpA enhances biofilm formation and PIA biosynthesis, whereas CodY represses PIA synthesis [57, 58]. In *S. epidermidis*, ethanol stress, iron limitation, and TCA cycle inactivation enhance biofilm formation and PIA synthesis [7, 13, 14, 16, 24, 59]. Based on the metabolomic data and published observations, it was reasonable to hypothesize that PIA biosynthesis was regulated in response to TCA cycle associated metabolomic changes by a CcpA-dependent and CodY-independent mechanism. To test this hypothesis, *ccpA* and *codY* deletion mutants were constructed in strains 1457 and 1457-*acnA*, and the amount of cell-associated PIA was determined after 6 h of growth (Figure 6.6). Consistent with previous observations (7), TCA cycle inactivation (strain 1457-*acnA*) dramatically increased the accumulation of PIA, whereas neither CodY nor CcpA had a dramatic effect on the post-exponential growth phase amount of PIA. When the *codY* mutation was introduced into an aconitase mutant background, PIA accumulation resembled the response in strain 1457-*acnA*, suggesting that TCA cycle-associated changes in PIA biosynthesis are independent of CodY. In contrast to the *codY-acnA* double mutant, the *ccpA-acnA* double mutant failed

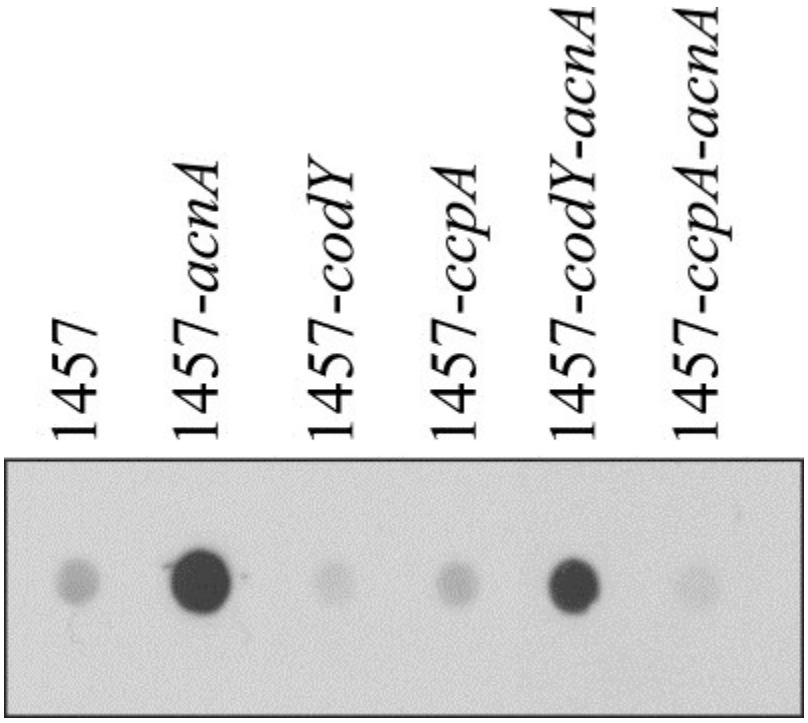


Figure 6.6. CcpA is required for PIA synthesis during TCA cycle stress. PIA immunoblot assay of strain 1457 and isogenic mutants of *acnA*, *codY*, *ccpA*, *codY*/*acnA*, and *ccpA*/*acnA* grown for 6 h inTSB. The results are representative of three independent experiments.

to produce PIA, strongly suggesting that some TCA cycle-associated metabolomics changes (*i.e.* glucose-6 phosphate) are sensed by CcpA, which in turn activates PIA biosynthesis.

6.3.6 TCA Cycle Stress Decreases RNAIII Transcription

In *S. aureus*, inactivation of the TCA cycle increases the transcription or stability of the riboregulator RNAIII of the accessory gene regulator (Agr) system [8, 9]. Conversely, increasing TCA cycle activity decreases the transcription or stability of RNAIII [60]. The causal relationship between TCA cycle activity and RNAIII transcript levels in *S. aureus* led us to examine whether disparate environmental conditions would similarly affect RNAIII transcription or message stability in *S. epidermidis*. In contrast to *S. aureus*, TCA cycle inactivation decreased RNAIII transcription or stability during the exponential and post-exponential growth phases in *S. epidermidis* (Figure 6.5). Importantly, ethanol stress and iron limitation decreased RNAIII transcription or stability in a similar manner to TCA cycle inactivation (Figure 6.5). In total, these data suggest that environmental stresses act through the TCA cycle to elicit transcriptional changes to at least two of the major staphylococcal virulence regulators (*i.e.* CcpA and RNAIII).

6.4 Discussion

In the life cycle of *S. epidermidis*, the transition from a skin resident, commensal state to adhering on implanted biomaterials represents a dramatic environmental change. In most pathogenic bacteria, environmental changes are accompanied by changes in the

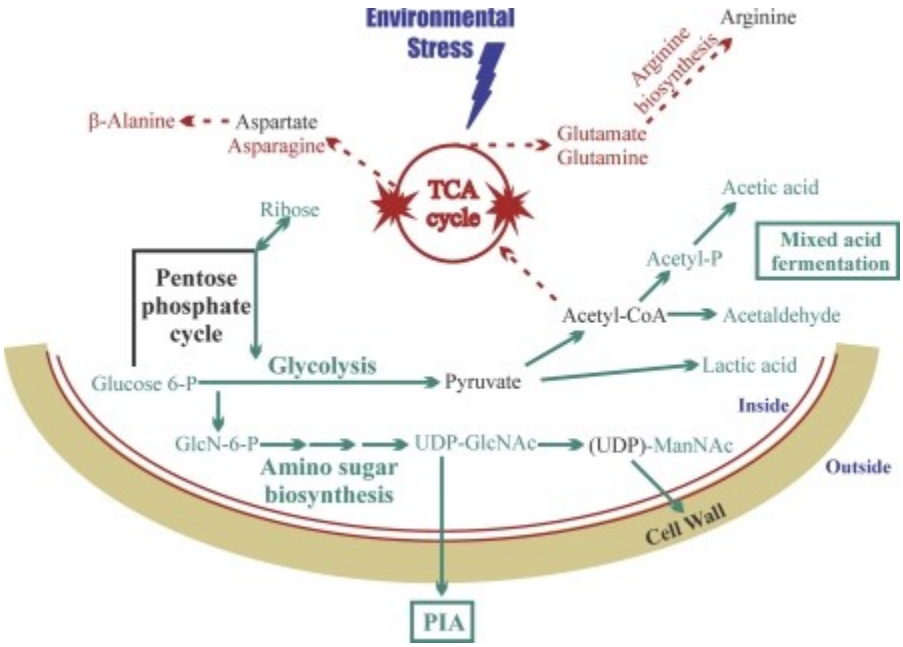


Figure 6.7. Summary of post-exponential growth phase metabolic changes associated with TCA cycle stress. Metabolites in *green* represent an increased concentration relative to the wild-type strain. Metabolites in *red* represent a decreased concentration relative to the wild-type strain. Metabolites and pathways in *black* are inferred from the data, but they are inconclusive. *Glucose-6-P*, glucose-6-phosphate; *GlcN-6-P*, GlcN-6-phosphate.

transcription of virulence genes; thus, environmental signals (*e.g.* nutrient replete, iron-limiting, or oxygen-limiting growth conditions) commonly regulate virulence gene transcription [2, 61-65]. Although *S. epidermidis* has relatively few virulence determinants, one of its primary pathogenic effectors is the exopolysaccharide PIA [66-69]. Previously, we demonstrated that PIA biosynthesis is regulated by TCA cycle activity; specifically, repression of TCA cycle activity dramatically enhances transcription of PIA biosynthetic genes (*icaADBC*) and PIA accumulation [7, 24, 60]. In this study, we demonstrate that dissimilar environmental signals decrease TCA cycle activity (Figure 6.1), resulting in common metabolomics changes (Figure 6.2 and Tables 6.1 and 6.2; summarized in Figure 6.7) that alter the activity of metabolite-responsive regulators such as CcpA (Figure 6.6). These data lead us to propose that it is the TCA cycle itself that is "sensing" the environmental transition and transducing this information into metabolic signals that activate or repress the activity of metabolite responsive regulators to modulate the expression of PIA and other virulence determinants.

As *S. epidermidis* transitions from residing on the skin to being implanted in a host, it enters into an environment where free iron is present at a concentration of 10^{-18} M [62], a condition antagonistic to TCA cycle activity (Figure 6.1) [9, 11]. Similarly, this transition dramatically decreases the availability of free oxygen (the partial pressure of atmospheric O_2 is 159mmHg at sea level, and this decreases to an estimated 3–5 mm Hg at the host cell level), a condition that is also antagonistic to TCA cycle activity. Taken together, this type of environmental transition is accompanied by conditions that are inhibitory to TCA cycle activity and stimulatory to PIA biosynthesis [7, 24]. In other

words, the transition from an external environment to an internal environment represses TCA cycle activity and primes *S. epidermidis* for PIA synthesis, which enhances biofilm formation and increases the likelihood of establishing a biomaterial-associated infection.

The significance of the data presented here is 5-fold. First, it establishes a mechanism by which well-established regulators (*e.g.* CcpA) participate in responding to environmental stresses. Second, these data suggest how disparate environmental stimuli can cause common phenotypic changes (*e.g.* iron limitation and ethanol stress both increase PIA synthesis and biofilm formation [13, 70]). Third, these data suggest that a difficulty in attributing the effects of an environmental stimulus, such as iron limitation, to a specific regulator, such as the ferric uptake regulator (Fur) [71], is that many of the effects are due to metabolite-responsive regulators reacting to changes in the metabolome. Fourth, *S. epidermidis* has a second general stress response system that is largely independent of the σ^B-controlled general stress response (Figure 6.2).

Finally, in bacteria, three metabolic pathways (Embden-Meyerhof-Parnas, pentose phosphate, and TCA cycle) produce the 13 biosynthetic intermediates needed to synthesize all macromolecules in a bacterial cell. Therefore, by linking virulence factor synthesis to the TCA cycle, bacteria are connecting virulence to the availability of biosynthetic intermediates needed to synthesize virulence determinants.

6.5 Conclusion

TCA cycle stress alters the intracellular concentrations of metabolites (Tables 6.1 and 5.2) relative to those of the wild-type strain 1457. If the change in the concentration

of a metabolite is sufficiently large, then the activity of a regulator that can respond to one of those metabolites may be altered. Therefore, these data present an exceptional opportunity to identify regulators that coordinate metabolism and virulence in *S. epidermidis*. Although a considerable amount of research needs to be done to determine which metabolite-responsive regulators are involved in responding to TCA cycle-associated metabolomics changes, the work presented here sheds light on how environmental signals alter the bacterial metabolic status to regulate adaptation to a new environment.

6.6 References

1. von Eiff, C., G. Peters, and C. Heilmann, *Pathogenesis of Infections Due to Coagulase-Negative Staphylococci.* The Lancet Infectious Diseases, 2002. **2**(11): p. 677-85.

2. Somerville, G.A. and R.A. Proctor, *At the Crossroads of Bacterial Metabolism and Virulence Factor Synthesis in Staphylococci.* Microbiology and Molecular Biology Reviews : MMBR, 2009. **73**(2): p. 233-48.

3. Begun, J., et al., *Staphylococcus aureus Virulence Factors Identified by Using a High-Throughput Caenorhabditis elegans-Killing Model.* Infection and Immunity, 2005. **73**(2): p. 872-7.

4. Coulter, S.N., et al., *Staphylococcus aureus Genetic Loci Impacting Growth and Survival in Multiple Infection Environments.* Molecular Microbiology, 1998. **30**(2): p. 393-404.

5. Mei, J.M., et al., *Identification of Staphylococcus aureus Virulence Genes in a Murine Model of Bacteraemia using Signature-Tagged Mutagenesis.* Molecular Microbiology, 1997. **26**(2): p. 399-407.

6. Bae, T., et al., *Staphylococcus aureus Virulence Genes Identified by Bursa aurealis Mutagenesis and Nematode Killing.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(33): p. 12312-7.

7. Sadykov, M.R., et al., *Tricarboxylic Acid Cycle-Dependent Regulation of Staphylococcus epidermidis Polysaccharide Intercellular Adhesin Synthesis.* Journal of Bacteriology, 2008. **190**(23): p. 7621-32.

8. Somerville, G.A., et al., *Staphylococcus aureus Aconitase Inactivation Unexpectedly Inhibits Post-Exponential-Phase Growth and Enhances Stationary-Phase Survival.* Infection and Immunity, 2002. **70**(11): p. 6373-82.

9. Somerville, G.A., et al., *Synthesis and Deformylation of Staphylococcus aureus delta-toxin are Linked to Tricarboxylic Acid Cycle Activity.* Journal of Bacteriology, 2003. **185**(22): p. 6686-94.

10. Collins, F.M. and J. Lascelles, *The Effect of Growth Conditions on Oxidative and Dehydrogenase Activity in Staphylococcus aureus.* Journal of General Microbiology, 1962. **29**: p. 531-5.

11. Somerville, G., C.A. Mikoryak, and L. Reitzer, *Physiological Characterization of Pseudomonas aeruginosa During Exotoxin A synthesis: Glutamate, Iron Limitation, and Aconitase Activity.* Journal of Bacteriology, 1999. **181**(4): p. 1072-8.

12. Varghese, S., Y. Tang, and J.A. Imlay, *Contrasting Sensitivities of Escherichia coli Aconitases A and B to Oxidation and Iron Depletion.* Journal of Bacteriology, 2003. **185**(1): p. 221-30.

13. Knobloch, J.K., et al., *Biofilm Formation by Staphylococcus epidermidis Depends on Functional RsbU, An Activator of the sigB Operon: Differential Activation Mechanisms Due to Ethanol and Salt Stress.* Journal of Bacteriology, 2001. **183**(8): p. 2624-33.

14. Presterl, E., et al., *Effects of Alcohols, Povidone-Iodine and Hydrogen Peroxide on Biofilms of Staphylococcus epidermidis.* The Journal of Antimicrobial Chemotherapy, 2007. **60**(2): p. 417-20.

15. Korem, M., et al., *Alcohol Increases Hemolysis by Staphylococci.* FEMS Microbiology Letters, 2007. **269**(1): p. 153-9.

16. Lyte, M., et al., *Stimulation of Staphylococcus epidermidis Growth and Biofilm Formation by Catecholamine Inotropes.* Lancet, 2003. **361**(9352): p. 130-5.

17. Otzen, D.E., P. Sehgal, and L.W. Nesgaard, *Alternative Membrane Protein Conformations in Alcohols.* Biochemistry, 2007. **46**(14): p. 4348-59.

18. Mack, D., N. Siemssen, and R. Laufs, *Parallel Induction by Glucose of Adherence and a Polysaccharide Antigen Specific for Plastic-Adherent Staphylococcus epidermidis: Evidence for Functional Relation to Intercellular Adhesion.* Infection and Immunity, 1992. **60**(5): p. 2048-57.

19. Handke, L.D., et al., *SigmaB and SarA Independently Regulate Polysaccharide Intercellular Adhesin Production in Staphylococcus epidermidis.* Canadian Journal of Microbiology, 2007. **53**(1): p. 82-91.

20. Horton, R.M., et al., *Gene Splicing by Overlap Extension: Tailor-Made Genes Using the Polymerase Chain Reaction.* BioTechniques, 1990. **8**(5): p. 528-35.

21. Kennedy, M.C., et al., *The Role of Iron in the Activation-Inactivation of Aconitase.* The Journal of Biological Chemistry, 1983. **258**(18): p. 11098-105.

22. Baughn, A.D. and M.H. Malamy, *A Mitochondrial-Like Aconitase in the Bacterium Bacteroides fragilis: Implications for the Evolution of the Mitochondrial Krebs Cycle.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(7): p. 4662-7.

23. Lowry, O.H., et al., *Protein Measurement with the Folin Phenol Reagent.* Journal of Biological Chemistry, 1951. **193**(1): p. 265-275.

24. Vuong, C., et al., *Staphylococcus epidermidis Polysaccharide Intercellular Adhesin Production Significantly Increases During Tricarboxylic Acid Cycle Stress.* Journal of Bacteriology, 2005. **187**(9): p. 2967-73.

25. Kay, L.E., P. Keifer, and T. Saarinen, *Pure Absorption Gradient Enhanced Heteronuclear Single Quantum Correlation Spectroscopy with Improved Sensitivity.* Journal of the American Chemical Society, 1992. **114**(26): p. 10663-10665.

26. Palmer, A.G., et al., *Sensitivity Improvement in Proton-Detected 2-Dimensional Heteronuclear Correlation Nmr-Spectroscopy.* Journal of Magnetic Resonance, 1991. **93**(1): p. 151-170.

27. Bax, A. and D.G. Davis, *Mlev-17-Based Two-Dimensional Homonuclear Magnetization Transfer Spectroscopy.* Journal of Magnetic Resonance, 1985. **65**(2): p. 355-360.

28. Forgue, P., et al., *NMR Metabolic Profiling of Aspergillus nidulans to Monitor Drug and Protein Activity.* Journal of Proteome Research, 2006. **5**(8): p. 1916-23.

29. Halouska, S., et al., *Use of NMR Metabolomics to Analyze the Targets of D-Cycloserine in Mycobacteria: Role of D-Alanine Racemase.* Journal of Proteome Research, 2007. **6**(12): p. 4608-14.

30. Halouska, S. and R. Powers, *Negative Impact of Noise on the Principal Component Analysis of NMR Data.* Journal of Magnetic Resonance, 2006. **178**(1): p. 88-95.

31. Nguyen, B.D., et al., *SOGGY: Solvent-Optimized Double Gradient Spectroscopy for Water Suppression. A Comparison With Some Existing Techniques.* Journal of Magnetic Resonance, 2007. **184**(2): p. 263-74.

32. Craig, A., et al., *Scaling and Normalization Effects in NMR Spectroscopic Metabonomic Data Sets.* Analytical Chemistry, 2006. **78**(7): p. 2262-7.

33. Piotto, M., V. Saudek, and V. Sklenar, *Gradient-Tailored Excitation for Single-Quantum NMR Spectroscopy of Aqueous Solutions.* Journal of Biomolecular NMR, 1992. **2**(6): p. 661-5.

34. Johnson, B.A., *Using NMRView to Visualize and Analyze the NMR Spectra of macromolecules.* Methods in Molecular Biology, 2004. **278**: p. 313-52.

35. Goddard, T.D.a.K., D.G., *Sparky 3.* University of California, San Francisco, Ca.

36. Cui, Q., et al., *Metabolite Identification via the Madison Metabolomics Consortium Database.* Nature Biotechnology, 2008. **26**(2): p. 162-4.

37. Ulrich, E.L., et al., *BioMagResBank.* Nucleic Acids Research, 2008. **36**(Database issue): p. D402-8.

38. Wishart, D.S., et al., *HMDB: the Human Metabolome Database.* Nucleic Acids Research, 2007. **35**(Database issue): p. D521-6.

39. Kanehisa, M., et al., *KEGG for Linking Genomes to Life and the Environment.* Nucleic Acids Research, 2008. **36**(Database issue): p. D480-4.

40. Karp, P.D., et al., *Expansion of the BioCyc Collection of Pathway/Genome Databases to 160 Genomes.* Nucleic Acids Research, 2005. **33**(19): p. 6083-9.

41. Werth, M.T., et al., *Analysis of Metabolomic PCA Data using Tree Diagrams.* Analytical Biochemistry, 2010. **399**(1): p. 58-63.

42. Efron, B., E. Halloran, and S. Holmes, *Bootstrap Confidence Levels for Phylogenetic Trees.* Proceedings of the National Academy of Sciences of the United States of America, 1996. **93**(23): p. 13429-34.

43. Felsenstein, J., *Confidence-Limits on Phylogenies - an Approach Using the Bootstrap.* Evolution, 1985. **39**(4): p. 783-791.

44. Retief, J.D., *Phylogenetic Analysis Using PHYLIP.* Methods in Molecular Biology, 2000. **132**: p. 243-58.

45. Somerville, G.A., et al., *Correlation of Acetate Catabolism and Growth Yield in Staphylococcus aureus: Implications for Host-Pathogen Interactions.* Infection and Immunity, 2003. **71**(8): p. 4724-32.

46. Wu, S., H. de Lencastre, and A. Tomasz, *Sigma-B, A Putative Operon Encoding Alternate Sigma Factor of Staphylococcus aureus RNA Polymerase: Molecular Cloning and DNA Sequencing.* Journal of Bacteriology, 1996. **178**(20): p. 6036-42.

47. Bischoff, M., et al., *Microarray-Based Analysis of the Staphylococcus aureus sigmaB Regulon.* Journal of Bacteriology, 2004. **186**(13): p. 4085-99.

48. Bischoff, M., J.M. Entenza, and P. Giachino, *Influence of a Functional sigB Operon on the Global Regulators sar and agr in Staphylococcus aureus.* Journal of Bacteriology, 2001. **183**(17): p. 5171-9.

49. Meier, S., et al., *sigmaB and the sigmaB-Dependent arlRS and yabJ-spoVG Loci Affect Capsule Formation in Staphylococcus aureus.* Infection and Immunity, 2007. **75**(9): p. 4562-71.

50. Pane-Farre, J., et al., *The sigmaB Regulon in Staphylococcus aureus and Its Regulation.* International Journal of Medical Microbiology : IJMM, 2006. **296**(4-5): p. 237-58.

51. Deutscher, J. and M.H. Saier, Jr., *ATP-Dependent Protein Kinase-Catalyzed Phosphorylation of a Seryl Residue in HPr, A Phosphate Carrier Protein of the Phosphotransferase System in Streptococcus pyogenes.* Proceedings of the National Academy of Sciences of the United States of America, 1983. **80**(22): p. 6790-4.

52. Bruckner, R. and F. Titgemeyer, *Carbon Catabolite Repression in Bacteria: Choice of the Carbon Source and Autoregulatory Limitation of Sugar Utilization.* FEMS Microbiology Letters, 2002. **209**(2): p. 141-8.

53. Sonenshein, A.L., *Control of Key Metabolic Intersections in Bacillus subtilis.* Nature Reviews. Microbiology, 2007. **5**(12): p. 917-27.

54. Warner, J.B. and J.S. Lolkema, *CcpA-Dependent Carbon Catabolite Repression in Bacteria.* Microbiology and Molecular Biology Reviews : MMBR, 2003. **67**(4): p. 475-90.

55. Shivers, R.P., S.S. Dineen, and A.L. Sonenshein, *Positive Regulation of Bacillus subtilis ackA by CodY and CcpA: Establishing a Potential Hierarchy in Carbon Flow.* Molecular Microbiology, 2006. **62**(3): p. 811-22.

56. Shivers, R.P. and A.L. Sonenshein, *Activation of the Bacillus subtilis Global Regulator CodY by Direct Interaction with Branched-Chain Amino Acids.* Molecular Microbiology, 2004. **53**(2): p. 599-611.

57. Seidl, K., et al., *Staphylococcus aureus CcpA Affects Biofilm Formation.* Infection and Immunity, 2008. **76**(5): p. 2044-50.

58. Majerczyk, C.D., et al., *Staphylococcus aureus CodY Negatively Regulates Virulence Gene Expression.* Journal of Bacteriology, 2008. **190**(7): p. 2257-65.

59. Matinaho, S., et al., *Dependence of Staphylococcus epidermidis on Non-Transferrin-Bound iron for Growth.* FEMS Microbiology Letters, 2001. **196**(2): p. 177-82.

60. Zhu, Y., et al., *Tricarboxylic Acid Cycle-Dependent Attenuation of Staphylococcus aureus in vivo Virulence by Selective Inhibition of Amino Acid Transport.* Infection and Immunity, 2009. **77**(10): p. 4256-64.

61. Bullen, J.J., H.J. Rogers, and E. Griffiths, *Role of Iron in Bacterial Infection.* Current Topics in Microbiology and Immunology, 1978. **80**: p. 1-35.

62. Litwin, C.M. and S.B. Calderwood, *Role of Iron in Regulation of Virulence Genes.* Clinical Microbiology Reviews, 1993. **6**(2): p. 137-49.

63. Mekalanos, J.J., *Environmental Signals Controlling Expression of Virulence Determinants in Bacteria.* Journal of Bacteriology, 1992. **174**(1): p. 1-7.

64. Park, M.K., R.A. Myers, and L. Marzella, *Oxygen Tensions and Infections: Modulation of Microbial Growth, Activity of Antimicrobial Agents, and Immunologic Responses.* Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America, 1992. **14**(3): p. 720-40.

65. Milenbachs, A.A., et al., *Carbon-Source Regulation of Virulence Gene Expression in Listeria monocytogenes.* Molecular Microbiology, 1997. **23**(5): p. 1075-85.

66. Vuong, C., et al., *A Crucial Role for Exopolysaccharide Modification in Bacterial Biofilm Formation, Immune Evasion, and Virulence.* The Journal of Biological Chemistry, 2004. **279**(52): p. 54881-6.

67. Vuong, C., et al., *Polysaccharide Intercellular Adhesin (PIA) Protects Staphylococcus epidermidis Against Major Components of the Human Innate Immune System.* Cellular Microbiology, 2004. **6**(3): p. 269-75.

68. Rupp, M.E., et al., *Characterization of Staphylococcus epidermidis Polysaccharide Intercellular Adhesin/Hemagglutinin in the Pathogenesis of Intravascular Catheter-Associated Infection in a Rat Model.* Infection and Immunity, 1999. **67**(5): p. 2656-9.

69. Rupp, M.E., *Infections of Intravascular Catheters and Vascular Devices.* In K. B. Crossley and G. L. Archer (ed.), The Staphylocci in Human Disease. Churchill-Livingstone, New York, NY, 1997: p. 379-399.

70. Deighton, M. and R. Borland, *Regulation of Slime Production in Staphylococcus epidermidis by Iron Limitation.* Infection and Immunity, 1993. **61**(10): p. 4473-9.

71. Torres, V.J., et al., *Staphylococcus aureus Fur Regulates the Expression of Virulence Factors That Contribute to the Pathogenesis of Pneumonia.* Infection and Immunity, 2010. **78**(4): p. 1618-28.

CHAPTER 7

NMR ANALYSIS OF A STRESS RESPONSE METABOLIC SIGNALING NETWORK[§]

7.1 Introduction

In eukaryotic organisms, signaling pathways are essential to the life-cycle of cells and are ubiquitous processes that regulate a variety of functions in response to both extracellular and intracellular environmental changes [1-4]. These signaling systems are spatially and temporarily organized, where the kinetic properties of these cycles depends on the cellular distribution of the activator and deactivator proteins. Protein activity is usually controlled through a variety of post-translational modifications (phosphorylation, acetylation, ubiquitylation, etc.), through protein complex formation, through transcription regulation, or any combination of these factors. A prototypical signaling cascade includes a membrane-bound receptor that binds a signal molecule which in turn activates a kinase proximal to the membrane. This activated kinase phosphorylates a second kinase, where the cascade continues and perpetuates the signal away from the membrane to the final target. Typically, the impact of a signaling network is the up- and down-regulation of a set of genes or proteins associated with a specific response (apoptosis, metabolic process, proliferation, stress responses, etc.). Correspondingly, the cell commits a significant amount of energy and resources to undergo such a phenotype change.

[§] Chapter 7 was adapted from Zhang, B., et. al., NMR Analysis of a Stress Response Metabolic Signaling Network, *Journal of Proteome Research* (2011), **10** (8), 3743-3754.

In prokaryotes, signal transduction frequently involves two-component regulatory systems that consist of a membrane-bound sensor histidine protein kinase and a response regulator [5]. These two-component signal transduction systems are activated when an external signaling molecule, peptide, metal ion, etc., is bound by the sensor kinase, which undergoes autophosphorylation at a conserved histidine. Transfer of the phosphoryl group to the receiver domain of a response regulator in the cytoplasm of a bacterium completes the activation and the response regulator is then competent to activate transcription of a limited set of genes. In staphylococci, there are numerous two-component regulatory systems [6], with the best studied being the *agr* quorum sensing system [7]. In addition to two-component systems, bacteria use sigma factors as a means to detect environmental conditions that induce heat stress, envelope stress, nitrogen stress, etc [8]. In staphylococci, σ^B is activated during stress conditions, growth phase transitions, and morphological changes [9, 10]. Most recently, the hypothesis that central metabolism can act as a signal transduction pathway to transduce external environmental signals (*e.g.*, iron-limitation) into intracellular metabolic signals by altering the activity of the enzymes of central metabolism has been proposed [11].

The tricarboxylic acid (TCA) cycle is part of central metabolism and provides reducing potential, energy and biosynthetic intermediates necessary for other macromolecular synthesis [12]. Several studies have also shown that the TCA cycle is involved in regulating or affecting virulence or virulent determinant biosynthesis [13-15]. One specific example is the production of the exopolysaccharide, polysaccharide intercellular adhesion (PIA) [16], which is associated with virulence and biofilm formation [17-20]. PIA synthesis is regulated by nutrient availability and external stress

conditions [21, 22]. Importantly, TCA cycle activity has also been shown to be affected by changes in environmental stress factors [16, 22, 23]. A number of environmental stress factors have also been shown to influence biofilm formation: ethanol [24], oleic acid [25], glucose [26], UDP-N-acetylglucosamine [27], sub-inhibitory concentrations of some antibiotics [28], anaerobic conditions [29], Fe limitation [30-32], high osmolarity [33], and high temperature [33]. The diversity of these external stimuli suggests a versatile regulation system. Recently, we used NMR metabolomics to demonstrate that Fe limitation and ethanol decrease TCA activity [23]. These stressors cause a common metabolic change that can be sensed by metabolite responsive-regulators (*e.g.*, catabolite control protein A; CcpA) that affect PIA production. We proposed that the TCA cycle plays a central role in a metabolic signaling network that senses disparate environmental stress conditions and regulates PIA biosynthesis, virulence determinants and biofilm formation (Figure 7.1). Herein, we report a further NMR analysis of the impact on the metabolome of *S. epidermidis* resulting from a diverse range of environmental stress factors associated with biofilm formation that include 5% NaCl [33], 2% glucose [26], 0.06 μg/mL tetracycline [28], and 400 nM autoinducer-2 (AI-2, furanosyl borate diester) [34], in addition to our prior study [23] with 4% ethanol [24], and Fe limitation [30-32].

Our differential NMR metabolomics methodology has been applied to the study of *in vivo* drug activity in *Aspergillus nidulans* [35] and *Mycobacterium smegmatis* [36]

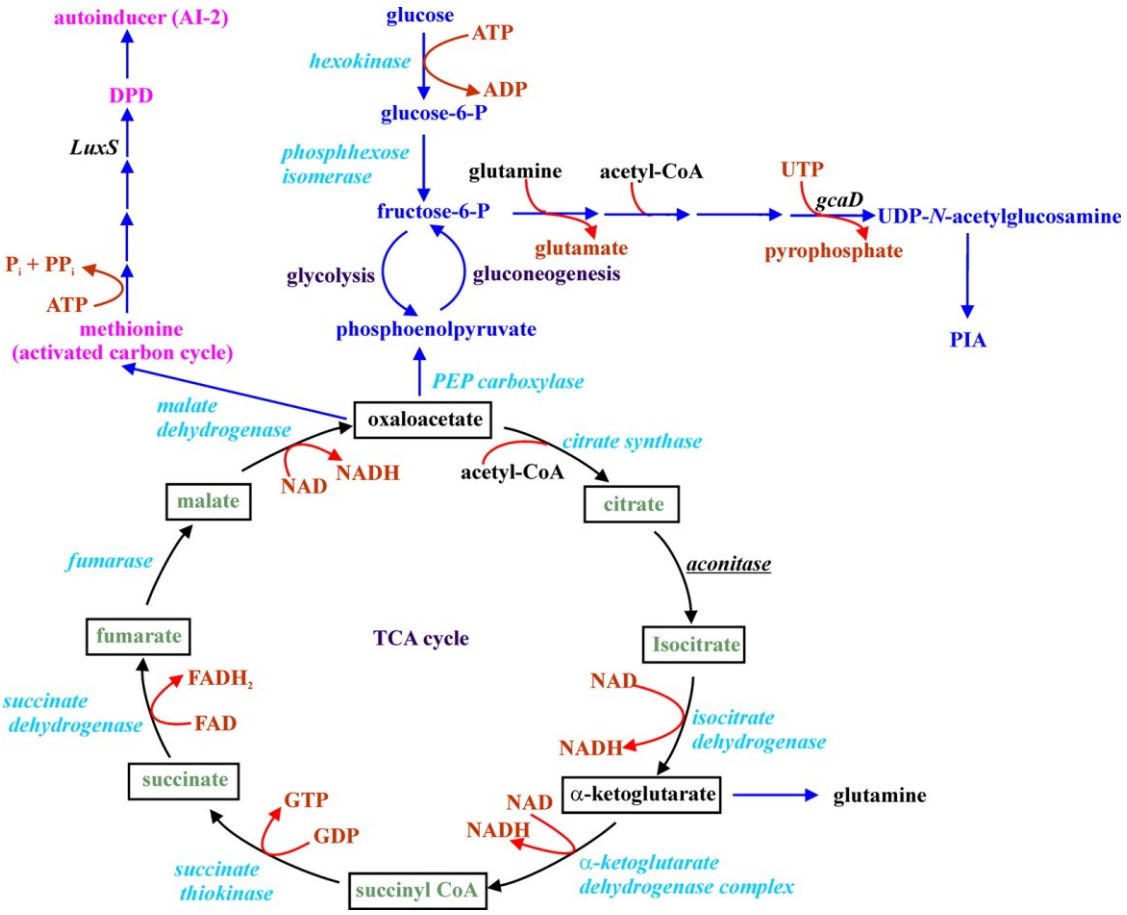


Figure 7.1. An illustration of the interrelationship of metabolic pathways associated with the TCA cycle and biofilm formation.

and is ideally suited to a systems biology analysis of the impact of environmental stress factors on the *S. epidermidis* metabolome and the corresponding role of the TCA cycle [23]. In the latter study, NMR was used to detect metabolic perturbations by following changes to *S. epidermidis* (strain 1457) cultured under environmental stress conditions that induce biofilm formation. These results were then compared to the metabolome of a *S. epidermidis* mutant (aconitase mutant strain SE1457-*acnA::tetM*) with an inactivated TCA cycle. If *S. epidermidis* senses environmental stress conditions by affecting TCA cycle activity as previously observed, then the impact on the metabolome caused by the aconitase mutant and the disparate external signals were expected to be equivalent. The overlapping clusters in principal component analysis (PCA) and orthogonal partial least squares discriminant analysis (OPLS-DA) two dimensional (2D) scores plot and the branch similarity on a metabolic tree diagram [37] indicate that external biofilm signals inactivate the TCA cycle. Furthermore, a detailed analysis of the relative concentration changes of 55 different metabolites from 2D ^{1}H-^{1}H TOCSY and 2D ^{1}H-^{13}C HSQC spectra implies the TCA cycle plays a central role in a metabolic signaling network [35, 36, 38, 39]. A metabolic network created with Cytoscape [40] illustrates this metabolic signaling network and the interrelationship of the TCA cycle activity with alanine metabolism, amino sugar metabolism, glycolysis/gluconeogenesis and the urea cycle.

The investigation of the NMR Analysis of a stress response metabolic signaling method was a group effort where Steven Halouska and Bo Zhang demonstrated significant and equal contribution to the project. Steven Halouska was involved in developing protocols to identify and quantify carbon-13 labeled metabolites derived from a 13C-Glucose. Bo Zhang performed a comparative analysis of metabolite fingerprints to

show the relationships of the impact between different environmental stress factors. Both Steven Halouska and Bo Zhang contributed equally in the analysis of the different stress responses.

7.2 Methods and Materials

7.2.1 Bacterial growth and NMR sample preparation.

Staphylococcus epidermidis wild-type strain 1457 and the isogenic aconitase mutant strain 1457-*acnA*::*tet*M were grown in tryptic soy broth (TSB; BD Biosciences) without dextrose and supplemented with 0.25% glucose (Sigma Chemical) or 0.25% $^{13}C_6$-glucose (Cambridge Isotope Laboratories).

All bacterial cultures were inoculated to an optical density at 600 nm ($O.D._{600}$) of 0.06 and were grown for 2 hours or 6 hours at 37^{o}C with 225 rpm aeration in TSB or TSB supplemented with a stressor know to induce biofilm formation. Either 10 or 12 replicate bacterial cultures were grown for each bacterial strain or environmental condition for the one-dimensional (1D) ^{1}H NMR experiments. 3 replicate bacterial cultures were grown for each bacterial strain or environmental condition for both the two-dimensional (2D) ^{1}H-^{13}C HSQC and the 2D ^{1}H-^{1}H TOCSY experiments.

In general, four different bacterial cultures were harvested per experiment: (1) wild-type *S. epidermidis* in TSB media, (2) wild-type *S. epidermidis* in TSB media with an environmental stress condition, (3) aconitase mutant of *S. epidermidis* in TSB, and (4) aconitase mutant of *S. epidermidis* in TSB with an environmental stress condition. The environmental stress conditions used in this study were 5% NaCl [33], 2% glucose [26], 0.06 μg/mL tetracycline [28], 400 nM autoinducer-2 (AI-2, furanosyl borate diester) [34]

in TSB media. To facilitate integrating this current work with our recent study [23] control cultures containing 4% ethanol [24] in TSB or TSB medium depleted of iron (DTSB, defferated TSB) [30-32] were included for comparison. DTSB was prepared as described [23]. AI-2, furanosyl borate diester, was synthesized as previously described [41, 42].

For the 1D ^{1}H NMR experiments, 2.74 $O.D._{600}$ units were harvested for analysis and for both the 2D ^{1}H-^{13}C HSQC and 2D ^{1}H-^{1}H TOCSY experiments 5.48 $O.D._{600}$ units were harvested. Following harvest, the culture medium was removed and the bacteria were suspended in 1 mL portions of 50 mM phosphate buffer (PBS) in 99.8% D_2O (Isotec) at pH 7.2 (uncorrected). The bacteria were lysed using a FAST-Prep instrument (MP Biomedicals) for 40 seconds, centrifuged for 5 min to remove the bacterial debris and glass beads, and frozen in liquid nitrogen.

7.2.2 NMR Data Collection.

The NMR spectra were collected on a Bruker 500 MHz Avance spectrometer equipped with a triple-resonance, Z-axis gradient cryoprobe. A BACS-120 sample changer with Bruker Icon software was used to automate the NMR data collection. 1D ^{1}H NMR spectra were collected using excitation sculpting [43] to efficiently remove the solvent and maintain a flat baseline, eliminating any need for baseline collection that may induce artifacts in the PCA or OPLS-DA 2D or three-dimensional (3D) scores plot. 1D ^{1}H NMR spectra were collected at 298K with a spectrum width of 5482.5 Hz and 32K data points. A total of 16 dummy scans and 64 scans were used to obtain each spectrum.

2D ^{1}H-^{13}C HSQC spectra were collected with solvent pre-saturation and relaxation delay of 0.5 seconds [44, 45]. A total of 1024 data points with a spectrum width of 4734.85 Hz, and 64 data points with a spectrum width of 13834.26 Hz were collected in the ^{1}H and ^{13}C dimensions, respectively. A total of 8 dummy scans and 128 scans were used to obtain each of the 2D ^{1}H-^{13}C HSQC NMR spectra. 2D ^{1}H-^{1}H TOCSY spectra were collected with WATERGATE solvent pre-saturation, and a relaxation delay of 2 seconds [46, 47]. A total of 1024 data points with a spectrum width of 5000 Hz, and 256 data points with a spectrum width of 5001.324 Hz were collected in the direct and indirect ^{1}H dimensions, respectively. A total of 16 dummy scans and 8 acquisition scans were used to obtain each of the 2D ^{1}H-^{1}H TOCSY NMR spectra.

7.2.3 NMR Data Analysis.

1D ^{1}H NMR spectra were processed in the ACD/1D NMR manager version 12.0 (Advanced Chemistry Development, Inc). The residual H_2O NMR resonance was removed. Intelligent bucketing was used to integrate each region with a bin size of 0.025 ppm. Each NMR spectrum was center averaged to minimize any experimental variations between bacterial cultures as follows:

$$Z = \frac{X_i - \bar{X}}{\sigma} \tag{6.1}$$

where $\bar{X}$ is the average signal intensity, σ is the standard deviation in the signal intensity, and X_i is the signal intensity within a bin. Noise regions of the spectra were omitted from the PCA analysis by setting the corresponding bins to zero [47]. The table of integrals was imported into SIMCA11.0+ (UMETRICS) for PCA and OPLS-DA analysis using

the program's standard parameters. The identification of the control group and treated group or groups for the OPLS-DA analysis was based on the PCA clustering pattern.

2D NMR Spectra were analyzed using NMRView [48] and Sparky (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco) to identify chemical shifts and assign peak intensities. Peak intensities were normalized for each 2D NMR spectrum by dividing by the average peak intensity for a given spectrum. Each peak for each metabolite from each specific triplicate data set was averaged and the intensity for each peak was further normalized across all data sets (i.e., wild-type, aconitase mutant, and each bacterial growth condition). Specifically, the maximal intensity for each peak across all data sets was set to 100. The peak intensities in the remaining data sets were all scaled relative to this peak intensity. Then, a normalized intensity for the metabolite within each data set was calculated by averaging the normalized intensity for each of the metabolite's assigned peaks. In this manner, the relative percent difference in metabolite intensity (concentration) can be reported between different bacterial strains or bacterial growth conditions. As an illustrative example, consider a metabolite with thee assigned peaks (*A, B, C*) in a 2D ^{1}H-^{13}C HSQC spectrum. The 2D ^{1}H-^{13}C HSQC spectrum is collected in triplicate under three different bacterial growth conditions for a total of 9 spectra and 27 peak intensities, 3 peaks in each spectrum for the metabolite. Peak *A* has average intensities of 0.05, 0.10, and 0.20 in the three different bacterial growth conditions, respectively. The values would be normalized to 25, 50, and 100. Similarly, peaks *B* and *C* are normalized against their maximal peak intensities for values of 20, 60, 100 and 30, 65, 100, respectively. Thus, the average relative concentrations of the metabolite under the three bacterial growth

conditions would be the average of the three normalized peaks, yielding values of 25, 58.3 and 100, respectively. The bacterial growth condition with the highest relative metabolite concentration (100) would have a corresponding concentration increase of 75 and 41.7 relative to the two other bacterial growth conditions.

The observed NMR peaks in the 2D ^{1}H-^{13}C HSQC and 2D ^{1}H-^{1}H TOCSY spectra were assigned to specific metabolites using ^{1}H and ^{13}C chemical shift tolerances of 0.05 ppm and 0.50 ppm, respectively. Metabominer [49], Madison Metabolomics Consortium Database (MMCD) [50], the BioMagResBank [51], and Human Metabolome Database [52] were used to identify metabolites. The presence of metabolites and metabolic pathways were verified with the KEGG [53] and Metacyc [54] databases. A metabolic network map was generated using Cytoscape using a force directed layout [40]. Metabolites identified with a percent concentration difference of $\geq \pm 10\%$ relative to wild-type *S. epidermidis* were manually color-coded to indicate either an up- or down-regulated concentration change.

7.3 Results and Discussion

7.3.1 NMR metabolomics and principle component analysis.

The elimination of experimental factors that may inadvertently influence PCA or OPLS-DA metabolomic data interpretation is essential. The observed variability in the PCA or OPLS-DA data should result from differences in the biological samples as opposed to changes in sample preparation, sample handling, data acquisitions, data processing or any number of experimental parameters (temperature, pH, time, concentration, etc.). In order to obtain accurate and reproducible metabolomic data, the

following general protocols were employed: (i) bacterial cultures of wild-type *S. epidermidis* were used as a reference metabolome and were prepared with all sets of bacterial cultures, (ii) equivalent bacterial numbers were used, so metabolite concentrations were independent of any bacterial growth variability, (iii) all NMR spectra were normalized using center averaging [55], so variability in sample concentration was minimized, (iv) noise regions [47] and solvents were removed from NMR spectra prior to PCA or OPLS-DA, and (v) minimal processing (no baseline correction or apodization functions) of the NMR spectra.

7.3.2 Harvesting of S. epidermidis cultures.

Bacteria grown *in vitro* undergo four different growth phases: (i) lag, (ii), exponential, (iii) stationary, and (iv) death. Throughout a typical growth cycle, the state of the bacteria and the environment are constantly changing. Clearly, there is a fundamental difference between the exponential phase, when cell density is relatively low, cells are rapidly dividing and the required nutrients are abundant; and the stationary phase when these characteristics are effectively reversed. Correspondingly, the metabolome is expected to reflect these differences. Therefore, exploring a biological system by monitoring changes in the metabolome necessitates the appropriate choice of the state of the system. In the case of *S. epidermidis* biofilm formation and the proposed role of the TCA cycle in a metabolic signaling network, the proper choice of the state of the system requires endogenous TCA cycle activity. The TCA cycle is minimally active during the exponential phase (2 h growth) when nutrients (*i.e.*, glucose or other rapidly catabolizable carbohydrates) are sufficient for bacteria to grow quickly [16, 56].

Overflow metabolism results in an incomplete oxidation of glucose, leading to the accumulation of acetate, lactate, and other incompletely oxidized metabolites in the culture medium (Figure 7.2a). During the transition to the post-exponential growth phase (6 hours), the TCA cycle is de-repressed as the carbohydrate(s) are depleted from the culture medium. Concomitantly, the incompletely oxidized metabolites that accumulated in the medium are catabolized through the TCA cycle resulting in the depletion of secondary metabolites from the culture medium.

The growth phase-dependent activity of the TCA cycle in *S. epidermidis* is also apparent from the PCA 2D scores plot generated from 1D ^{1}H NMR spectra of *S. epidermidis* cell lysate (Figure 7.2b). The 2D PCA scores plot indicates that PC1 and PC2 account for 20.6% and 10.7% of the variations in the NMR spectra, respectively. A 3D PCA scores plot (Figure 7.3) did not improve cluster separations. Each 1D ^{1}H NMR spectrum obtained for each cell lysate is represented as a single point in the PCA 2D scores plot, where the 10 replicates form four distinct clusters for the wild-type and aconitase mutant strains grown for 2 hours and 6 hours, respectively. As expected and consistent with our prior study [23], the metabolomes of the *S. epidermidis* wild-type and aconitase mutant cells from the exponential growth phase (2 hours) were more similar to each other than the 6 hour cultures. This is apparent from the close clustering in the 2D scores plot for the 2 hour wild-type and aconitase mutant. This is consistent with the minimal activity of the TCA cycle at 2 hours and the loss of TCA cycle activity for the

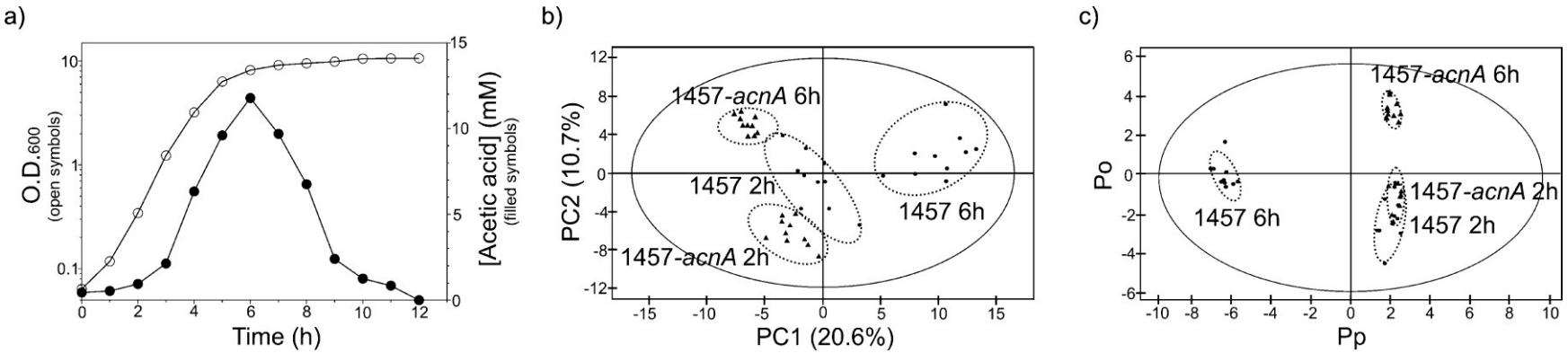


Figure 7.2. a) A typical *S. epidermidis* growth curve superimposed on the cellular production of acetic acid. **b)** 2D PCA scores plot and **c)** 2D OPLS-DA scores plot comparing 2 h growth of wild-type *S. epidermidis* 1457 (●), 2 h growth of aconitase mutant strain 1457-*acnA*::*tet*M (▲), 6 h growth of wild-type *S. epidermidis* 1457 (●), and 6 h growth of aconitase mutant strain 1457-*acnA*::*tet*M (▲). The ellipses correspond to the 95% confidence limits from a normal distribution for each cluster. For the OPLS-DA scores plot, the 6 h growth of wild-type *S. epidermidis* 1457 (●) was designated the control class and the remainder of the cells were designated as treated. The OPLS-DA used 1 predictive component and 3 orthogonal components to yield a R^2X of 0.788, R^2Y of 0.992 and Q^2 of 0.992.

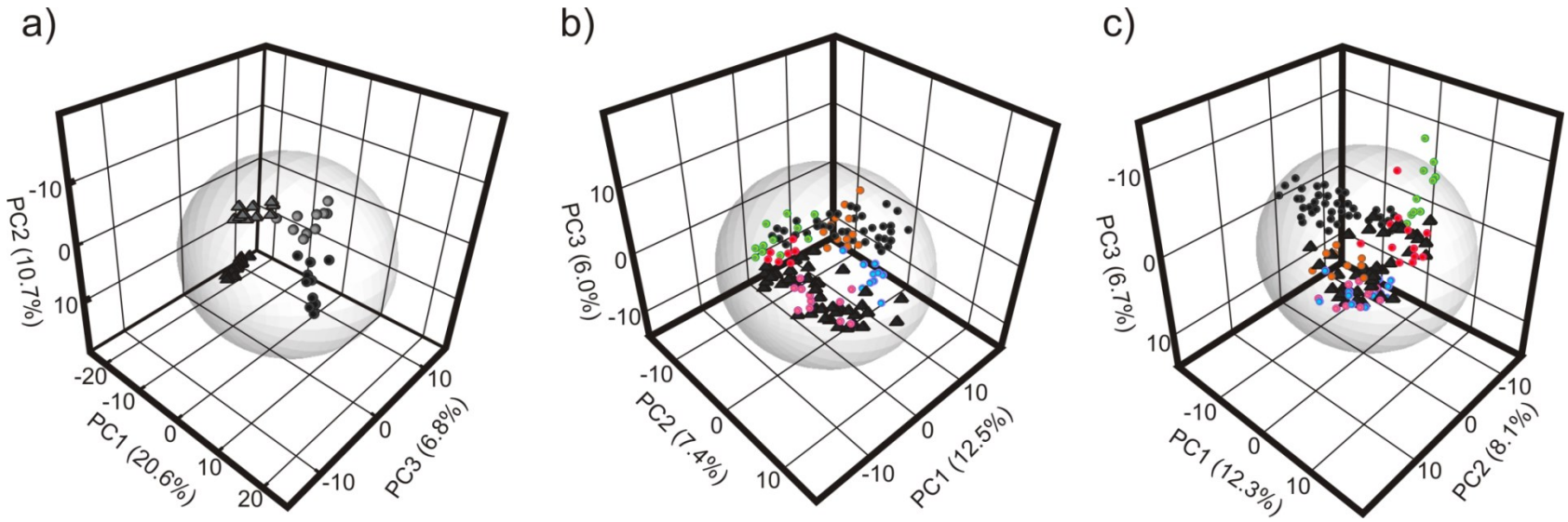

a)
PC2 (10.7%)
-10
0
10
-20
-10
0
10
20
PC1 (20.6%)
10
0
-10
PC3 (6.8%)
b)
PC3 (6.0%)
10
0
-10
-10
0
10
PC2 (7.4%)
10
0
-10
PC1 (12.5%)
c)
PC3 (6.7%)
-10
0
10
-10
0
10
PC1 (12.3%)
10
0
-10
PC2 (8.1%)

Figure 7.3. Three dimensional versions of the 2D PCA scores presented in Figure 7.2b, Figure 7.5b, and Figure 7.7b, respectively. **a)** 3D PCA scores plot comparing *S. epidermis* 1457 strain (circles) and aconitase mutant (triangles) grown for 2 hours (grey) and 6 hours (black). The PCA model consists of 3 significant components where the contribution (R^2) for each successive component is 20.6%, 10.7%, and 6.8%. The overall cross validation (Q^2) is 15.9%, 7.58%, and 2.7%, respectively. **b)** 3D PCA scores plot comparing wild-type *S. epidermidis* 1457 cells grown 6 h in standard TSB media (●), with *S. epidermidis* 1457 cells grown 6 h in iron-depleted media (DTSB) (●), with the addition of 4% ethanol (●),with the addition of 2% glucose (●), with the addition of 0.06 µg/mL tetracycline (●), with the addition of 5% NaCl (●), and 6 h growth of aconitase mutant strain 1457-*acnA*::*tet*M in standard TSB media (▲). The PCA model consists of 5 significant components where the contribution (R^2) for each successive component is 12.5%, 7.5%, 6.0%, 3.8%, and 3.3%. The overall cross validation (Q^2) is 15.9%, 7.58%, and 2.7%, respectively. **c)** 3D PCA scores plot comparing wild-type *S. epidermidis* 1457 cells grown 6 h in standard TSB media (●), 6 h growth of aconitase mutant strain 1457-*acnA*::*tet*M in standard TSB media (▲), aconitase mutant strain 1457-*acnA*::*tet*M in iron-depleted media (DTSB) (●), with the addition of 4% ethanol (●), with the addition of 2% glucose (●), with the addition of 0.06 µg/mL tetracycline (●), and with the addition of 5% NaCl (●). The PCA model consists of 5 significant components where the contribution (R^2) for each successive component is 12.3%, 8.1%, 6.7%, 3.8%, and 3.4%. The overall cross validation (Q^2) is 10.1%, 6.8%, 6.0, 1.6%, and 2.3%, respectively.

aconitase mutant. Conversely, there is a large separation in the 2D scores plot along the PC1 axis between the 6 hour wild-type and aconitase mutant. In fact, the 6 hour aconitase mutant cluster is closer to the 2 hour wild-type cluster. Again, this is consistent with an increase in TCA activity at 6 hours and the loss of TCA activity in the aconitase mutant. Correspondingly, the separation along PC1 reflects TCA cycle activity. Since the TCA cycle is minimally active during the exponential phase, the 2 hour wild-type cluster is slightly closer to the 6 hour wild-type cluster along PC1, compared to the aconitase mutants.

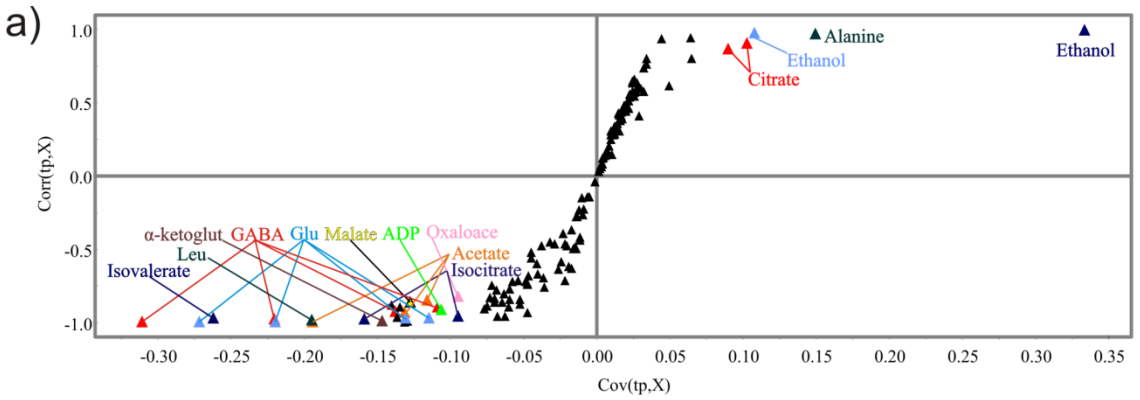
Alternatively, the separation along PC2 axis may reflect the variability in nutrients available to the cells. Glucose is still present after 2 hours of bacterial growth, but is being depleted while acetate is being accumulated. After 6 hours of growth, the depletion of acetate is dependent on TCA cycle activity, resulting in the largest separation along PC2 between the 6 hour aconitase mutant and the 2 hour wild-type samples. These two samples correspond to the largest expected variation in glucose and acetate concentrations. From the detailed 2D NMR analysis (*please see below*), acetate is approximately twice as concentrated in the 6 hour wild-type sample compared to the 2 hour aconitase mutant. The glucose concentration is effectively reversed. Glucose is twice as concentrated in the 2 hour aconitase mutant compared to the 6 hour wild type sample.

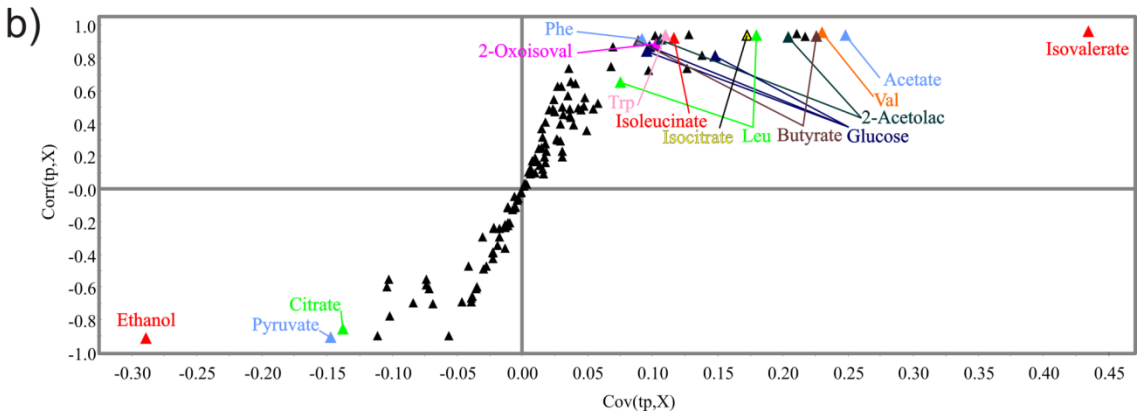
The PCA results were used to guide a subsequent analysis using OPLS-DA (Figure 7.2c). The OPLS-DA yielded a reliable model (R^2X 0.788, R^2Y 0.992, Q^2 0.992). The R^2 and Q^2 values represent the goodness of fit and predictability of the model, respectively. The OPLS-DA scores plot is similar to the PCA scores plot, except for the

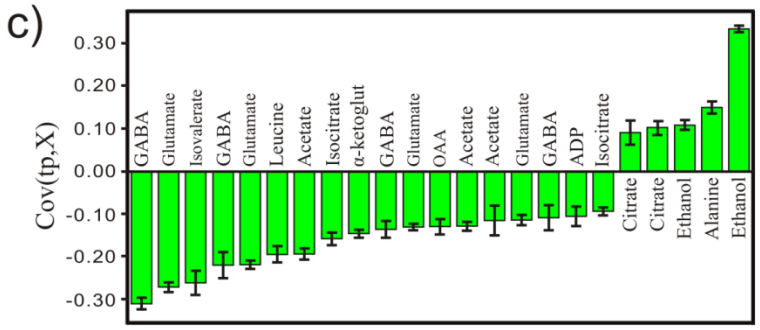
limited separation between the 2 hour wild-type and aconitase mutant clusters. Again, this is consistent with the minimal TCA activity expected for the 2 hour wild-type and aconitase mutant. Also, OPLS-DA emphasizes the difference between the control group (6 hour wild-type) and the treated classes, while minimizing contributions from within group variations. Thus, OPLS-DA generates significantly tighter clusters than PCA. More importantly, the S-plot (Figure 7.4) generated from the OPLS-DA provides unambiguous identification of the major contributors to the class separation (i.e., ^{1}H NMR bins and associated metabolites). Specifically, comparing the 6 hour aconitase mutant and the 6 hour wild-type samples, which had the largest separation along PC1 in the PCA 2D scores plot, identified citrate, isocitrate, and other TCA related metabolites. Similarly, comparing the 2 hour aconitase mutant and the 2 hour wild-type samples, which had the largest separation along PC2 in the PCA 2D scores plot, identified glucose, acetate and other nutrients required for cell growth. These results provide strong support for our subjective analysis of the trends in the 2D PCA scores plot and demonstrate that not only does PCA and OPLS-DA differentiate between metabolic profiles, but they also provide information about specific enzymatic activity and environmental conditions.

7.3.3 Impact of environmental stress conditions on the S. epidermidis metabolome.

Numerous genes and protein complexes are involved in the transformation of planktonic cells to a biofilm [57]. This process requires that *S. epidermidis* "sense" changes in its environment and the availability of nutrients, such as changes in temperature, O_2 levels, osmorality, ethanol, glucose and iron [29, 31, 33, 58-62]. It is reasonable to expect that different external factors would trigger distinct signaling


a)
Corr(tp,X)
Cov(tp,X)
Alanine
Ethanol
Citrate
Ethanol
α-ketoglut
GABA
Glu
Malate
ADP
Oxaloace
Leu
Isovalerate
Acetate
Isocitrate


b)
Corr(tp,X)
Cov(tp,X)
Phe
2-Oxoisoval
Isovalerate
Acetate
Val
Trp
Isoleucinate
2-Acetolac
Isocitrate
Leu
Butyrate
Glucose
Citrate
Ethanol
Pyruvate


c)
Cov(tp,X)
GABA
Glutamate
Isovalerate
GABA
Glutamate
Leucine
Acetate
Isocitrate
α-ketoglut
GABA
Glutamate
OAA
Acetate
Acetate
Glutamate
GABA
ADP
Isocitrate
Citrate
Citrate
Ethanol
Alanine
Ethanol

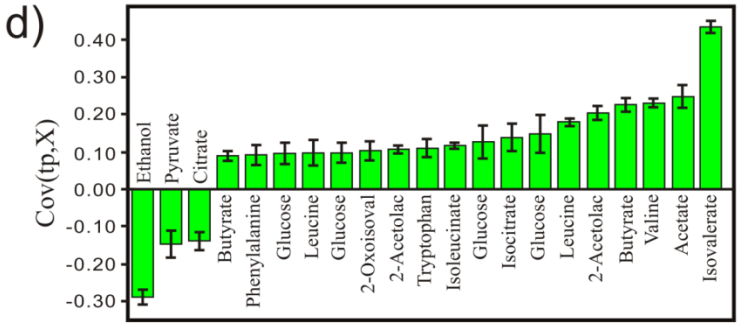

d)
Cov(tp,X)
Ethanol
Pyruvate
Citrate
Butyrate
Phenylalanine
Glucose
Leucine
Glucose
2-Oxoisoval
2-Acetolac
Tryptophan
Isoleucinate
Glucose
Isocitrate
Glucose
Leucine
2-Acetolac
Butyrate
Valine
Acetate
Isovalerate

Figure 7.4. a) OPLS-DA S-plots comparing the *S. epidermidis* 1457 and aconitase mutant strain 1457-*acnA*::*tet*M where both cell cultures were grown for 6 hours. The two cell cultures were shown be separated along the PC1 axis in Figure 7.2a. Each point in the S-plot represents a specific bin containing a chemical shift range of 0.025 ppm, where the points at the extreme ends of the S-plot are the major contributors to the class distinction. Each point was identified to a specific metabolite using the Human Metabolomics Database and Madison Metabolomics Database. All the identified metabolites are associated with TCA cycle inactivation. **b)** OPLS-DA S-plot comparing the mutant strain 1457-*acnA*::*tet*M grown for 2 hours and 6 hours. The two conditions are shown to be separated along the PC2 axis in Figure 7.2a. The metabolites identified are associated with variations in the utilization of glucose for cell growth. **c)** OPLS-DA loading plot comparing *S. epidermidis* 1457 and aconitase mutant strain 1457-*acnA*::*tet*M where both cell cultures were grown for 6 hours. Negative values indicate a decrease in peak intensity when comparing the wild type to the mutant, while positive values indicate an increase in peak intensity. These results are comparable to the bar graphs depicted in figure 7.9 from the analysis of 2D NMR data. **d)** OPLS-DA loading plot comparing the mutant strain 1457-*acnA*::*tet*M grown for 2 hours and 6 hours.

pathways and mechanisms of biofilm regulation. Correspondingly, different biofilm formation pathways would presumably induce dissimilar metabolomic profiles. Alternatively, a versatile regulation system responsive to disparate signals would be significantly more efficient. A metabolic signaling pathway is one potential mechanism of rapidly responding to changing environmental stress conditions. Conceptually, the environmental flux of essential nutrients and metabolites would direct the up- or down-regulation of specific metabolic pathways in response to concentration changes (Le Chatelier's principle) to initially reestablish equilibrium without affecting protein activity. Effectively, a limited or abundant metabolite would direct the metabolic flow through a specific pathway causing a cascade affect due to the high interrelationship of the metabolome. Eventually, gene and protein regulation processes would respond to the perturbed metabolic activity leading to the up- or down- regulation of specific genes and proteins.

We have previously demonstrated that both Fe limitation and 4% ethanol decrease TCA cycle activity [23]. These environmental stress factors are known to induce *S. epidermidis* biofilm formation [24, 30-32]. We have also demonstrated that Fe limitation and 4% ethanol had a similar impact on the *S. epidermidis* metabolome and altered the activity of CcpA, a metabolite-responsive regulator. An important role for the TCA cycle in a staphylococcal biofilm metabolic signaling pathway seems apparent, especially since the TCA cycle is a central metabolic pathway that interacts with numerous other pathways (Figure 7.1). Thus, we proposed that the TCA cycle senses environmental stressors and transduces this signal through the metabolome to activate or repress the activity of metabolite-responsive regulators, which, in turn, modulates PIA production,

virulence factor synthesis, and biofilm formation [23]. To further support our hypothesis that the TCA cycle senses disparate environmental signals to regulate PIA synthesis and biofilm formation, we analyzed changes in the *S. epidermidis* metabolome caused by additional environmental stress factors also known to induce an *S. epidermidis* biofilm [26, 28, 34, 63].

S. epidermidis cultures were treated with 5% NaCl, 2% glucose, 0.06 μg/mL tetracycline, and 400 nM of autoinducer-2 (AI-2, furanosyl borate diester). Glucose and NaCl were reported to induce biofilm formation by the regulation of the *rbf* gene [64], which has been shown to be a regulator of *icaR* [65], a negative regulator of the *icaADBC* operon that is required for PIA synthesis and biofilm formation [66]. Subinhibitory concentrations of antibiotics enhance *icaADBC* gene expression [28, 67] by potentially inhibiting TcaR, a weak negative regulator of *icaADBC* gene expression [68]. AI-2 is an intercellular signaling molecule that has a modest effect on staphylococcal biofilms [63]. Conversely, if these additional environmental stress factors impact the *S. epidermidis* metabolome in a manner similar to Fe limitation and 4% ethanol, which is also correlated with TCA cycle inactivation, then these results would further support the hypothesis that the TCA cycle acts as a signal transducer as a part of a metabolic signaling network.

The PCA 2D scores plot (Figure 7.5a) and the associated metabolic tree (Figure 7.5c) indicates that *S. epidermidis* wild-type cultures grown with the addition of the environmental stressors 4% ethanol, 0.06 μg/mL tetracycline or iron-limitation exhibited essentially identical metabolomes as the aconitase mutant. Both the aconitase mutant and wild-type cultures under these stress conditions formed a large cluster distinct from the

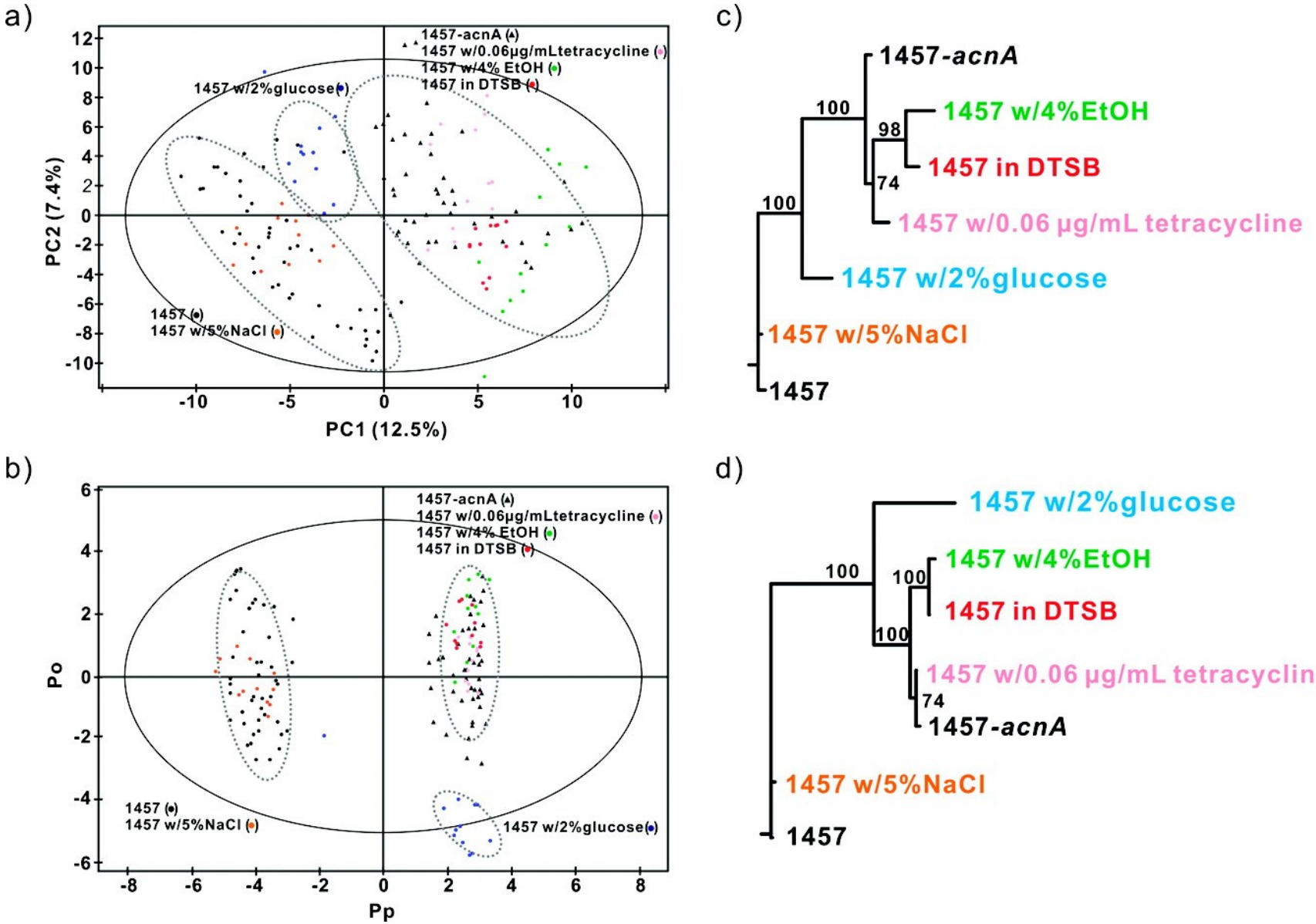

a)
1457-acnA (▲)
1457 w/0.06μg/mLtetracycline (•)
1457 w/4% EtOH (•)
1457 in DTSB (•)
1457 w/2%glucose (•)
1457 (•)
1457 w/5%NaCl (•)
PC2 (7.4%)
PC1 (12.5%)
b)
1457-acnA (▲)
1457 w/0.06μg/mLtetracycline (•)
1457 w/4% EtOH (•)
1457 in DTSB (•)
1457 (•)
1457 w/5%NaCl (•)
1457 w/2%glucose (•)
Po
Pp
c)
1457-acnA
1457 w/4%EtOH
1457 in DTSB
1457 w/0.06 μg/mL tetracycline
1457 w/2%glucose
1457 w/5%NaCl
1457
100
98
74
100
d)
1457 w/2%glucose
1457 w/4%EtOH
1457 in DTSB
1457 w/0.06 μg/mL tetracycline
1457-acnA
1457 w/5%NaCl
1457
100
100
100
74

Figure 7.5. **a)** 2D PCA scores plot and **b)** 2D OPLS-DA comparing wild-type *S. epidermidis* 1457 cells grown 6 h in standard TSB media (●), with *S. epidermidis* 1457 cells grown 6 h in iron-depleted media (DTSB) (●), with the addition of 4% ethanol (●),with the addition of 2% glucose (●), with the addition of 0.06 µg/mL tetracycline (●), with the addition of 5% NaCl (●), and 6 h growth of aconitase mutant strain 1457-*acnA*::*tet*M in standard TSB media (▲). The ellipses correspond to the 95% confidence limits from a normal distribution for each cluster. For the OPLS-DA scores plot, the 6 h growth of wild-type *S. epidermidis* 1457 (●) was designated the control class and the remainder of the cells were designated as treated. The OPLS-DA used 1 predictive component and 4 orthogonal components to yield a R^2X of 0.637, R^2Y of 0.966 and Q^2 of 0.941. Metabolomic tree diagram generated from the **c)** 2D PCA scores plot depicted in (a) and (d) 2D OPLS-DA scores plot depicted in (b). The label colors match the symbol colors from the 2D scores plot. Each node is labeled with the boot-strap number, where a value above 50 indicates a statistically significant separation.

cluster of wild-type cells in standard growth media. These results further support our hypothesis that environmental stress factors influence biofilm formation by inactivating the TCA cycle and re-directing key metabolites into PIA synthesis. Conversely, growing *S. epidermidis* wild-type cells in the presence of 5% NaCl showed no significant effect on the metabolome since the wild-type cells in the presence and absence of 5% NaCl cluster together. Similarly, AI-2 did not affect the *S. epidermidis* metabolome since both the wild-type and aconitase mutant cells in the presence and absence of 400 nM AI-2 cluster together (Figure 7.6). Interestingly, *S. epidermidis* cells treated with 2% glucose were separated from both the wild-type and aconitase mutant clusters, implying a different impact on the metabolome and a unique mechanism of regulating biofilm formation. Alternatively, the addition of 2% glucose may be viewed as an intermediary effect, where the metabolome of the *S. epidermidis* cells grown with 2% glucose is moving toward the aconitase mutant cells. Effectively, the amount of glucose added to the bacterial culture was insufficient to completely inactivate the TCA cycle. It has been previously shown that different strains have different glucose uptake rates and different sensitivities to glucose-induced biofilm formation [69, 70].

The PCA results were used to guide a subsequent analysis using OPLS-DA (Figure 7.5b) and the corresponding metabolomics tree diagram (Figure 7.5d). The OPLS-DA analysis yielded a reliable model (R^2X 0.637, R^2Y 0.966, Q^2 0.941), and results identical to PCA. The wild-type cells in the presence and absence of 5% NaCl were defined as the controls and, as expected, formed a single cluster in the 2D scores plot. *S. epidermidis* wild-type cultures grown with the addition of the environmental stressors 4% ethanol, 0.06 μg/mL tetracycline or iron-limitation again formed a single

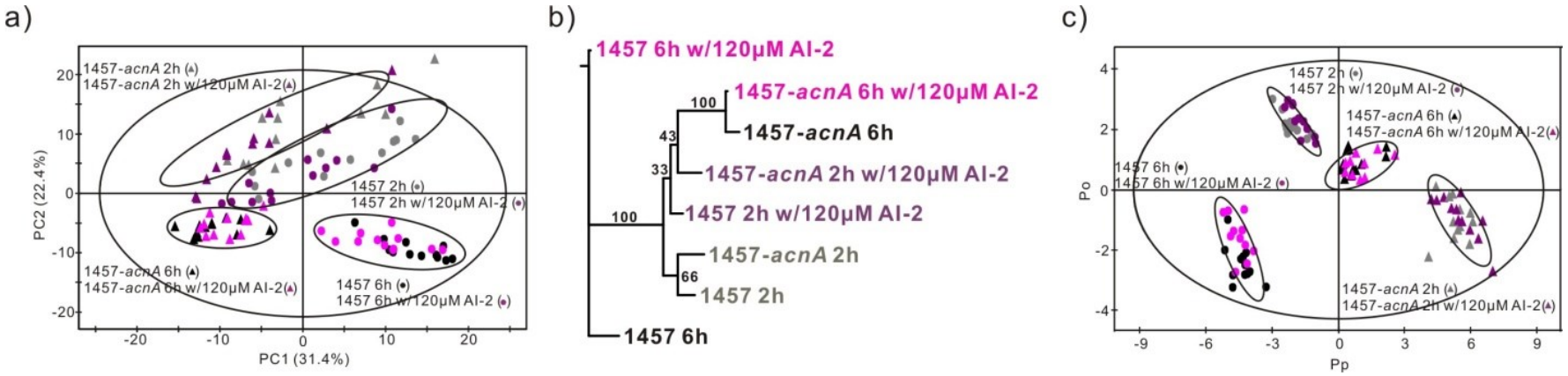
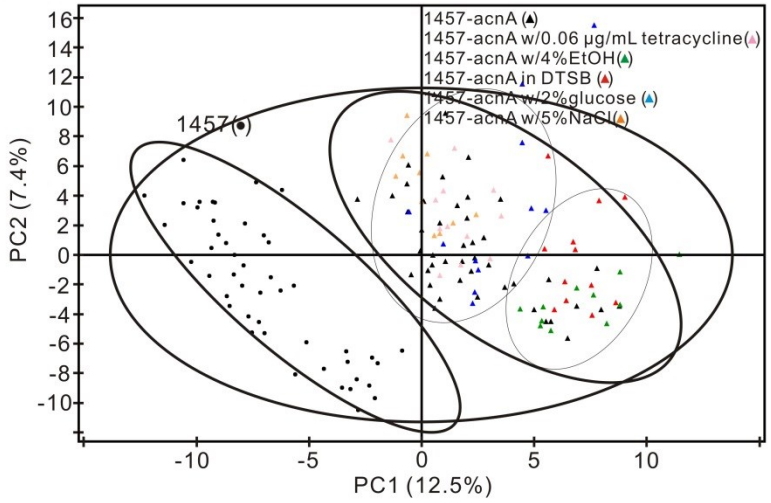


Figure 7.6. a) 2D PCA scores plot comparing wild-type *S. epidermidis* strain 1457 and aconitase mutant strain 1457-*acnA*::*tet*M with and without the addition of 400 nM of AI-2. Cells were grown for 2 h or 6 h in standard TSB media, 2 h growth of wild-type *S. epidermidis* strain 1457 with AI-2 (●) and without AI-2 (●); aconitase mutant strain 1457-*acnA*::*tet*M with AI-2 (▲) and without AI-2 (▲); 6 h growth of wild-type *S. epidermidis* strain 1457 with AI-2 (●) and without AI-2. (●); and aconitase mutant strain 1457-*acnA*::*tet*M with AI-2 (▲) and without AI-2 (▲). **b)** Metabolomic tree diagram generated from the 2D PCA scores plot depicted in (a). The label colors match the symbol colors from the 2D PCA scores plot. Each node is labeled with the boot-strap number, where a value above 50 indicates a statistically significant separation. **c)** 2D OPLS-DA scores plot comparing wild-type *S. epidermidis* strain 1457 and aconitase mutant strain 1457-*acnA*::*tet*M with and without the addition of 400 nM of AI-2. Each pair of cell cultures treated with and without AI-2 were defined as a separate class for a total of four separate classes. The model consists of 1 predictive component and 3 orthogonal components that yielded an R^2X of 0.774, R^2Y of 0.970, and a Q^2 of 0.954.

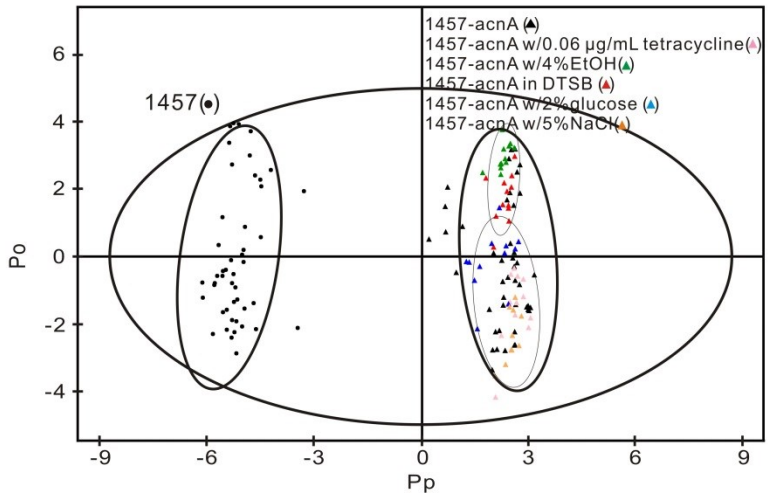
cluster with the aconitase mutant in the 2D OPLS-DA scores plot. Also similar to the PCA results, the *S. epidermidis* cells grown with 2% glucose formed a unique cluster. Thus, the corresponding metabolomics tree diagram identified three distinct clusters with bootstrap values of 100. The OPLS-DA results further support our hypothesis that environmental stress factors influence biofilm formation by inactivating the TCA cycle and re-directing key metabolites into PIA synthesis.

To verify the observed effect on the *S. epidermidis* metabolome is due to inactivating the TCA cycle as opposed to other potential factors, the *S. epidermidis* aconitase mutant strain was also grown with the addition of 4% ethanol, 0.06 μg/mL tetracycline, 2% glucose, 5% NaCl or under iron-limitation conditions. If the impact of these stress conditions is primarily through the inactivation of the TCA cycle, then the metabolome of the aconitase mutant strain should be unperturbed. Otherwise, if the stress conditions induce additional or alternative responses, then changes in the metabolome should be observed. The 2D PCA scores plot (Figure 7.7a) and metabolomic tree (Figure 7.7c) indicate the stress conditions did not affect the metabolome of the aconitase mutant. The aconitase mutant with and without the stress conditions forms a large cluster distinct from the wild-type cluster. Importantly, this includes the addition of 2% glucose. This implies the addition of 2% glucose to wild-type *S. epidermidis* resulted in an incomplete inactivation of the TCA cycle instead of a novel mechanism of biofilm regulation. Again, the PCA results were used to guide a subsequent analysis using OPLS-DA (Figure 7.7b) and the corresponding metabolomics tree diagram (Figure 7.7d). The OPLS-DA analysis yielded an acceptable model (R^2X 0.488, R^2Y 0.976, Q^2 0.961), and results very similar to PCA. The lower R^2X is consistent with the larger spread observed within the two

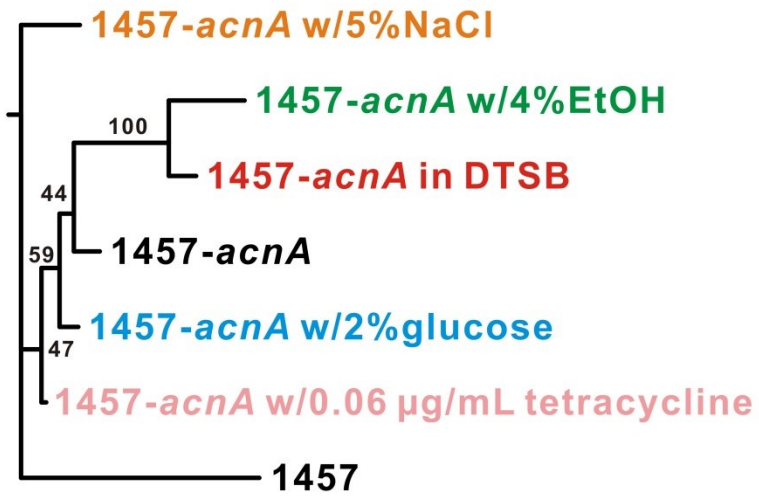
a)


1457-acnA (▲)
1457-acnA w/0.06 μg/mL tetracycline(▲)
1457-acnA w/4%EtOH(▲)
1457-acnA in DTSB (▲)
1457-acnA w/2%glucose (▲)
1457-acnA w/5%NaCl(▲)
1457(●)
PC2 (7.4%)
PC1 (12.5%)

b)


1457-acnA (▲)
1457-acnA w/0.06 μg/mL tetracycline(▲)
1457-acnA w/4%EtOH(▲)
1457-acnA in DTSB (▲)
1457-acnA w/2%glucose (▲)
1457-acnA w/5%NaCl(▲)
1457(●)
Po
Pp

c)


1457-acnA w/5%NaCl
1457-acnA w/4%EtOH
100
1457-acnA in DTSB
44
1457-acnA
59
1457-acnA w/2%glucose
47
1457-acnA w/0.06 μg/mL tetracycline
1457

d)

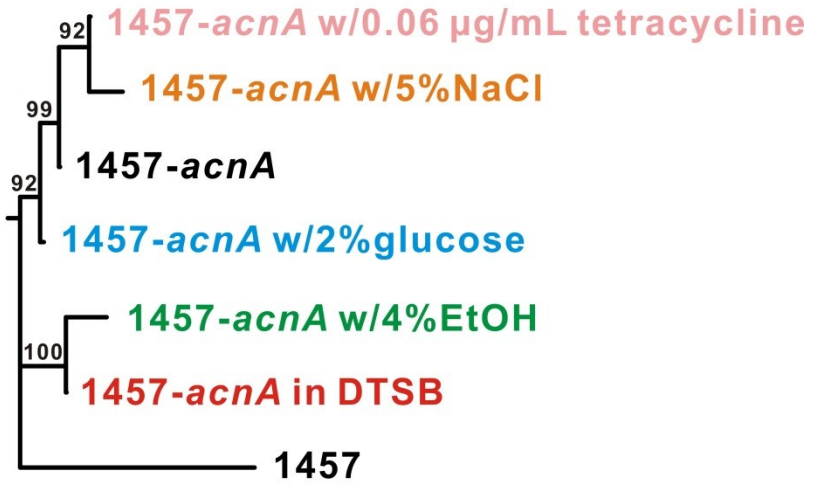

92
1457-acnA w/0.06 μg/mL tetracycline
1457-acnA w/5%NaCl
99
1457-acnA
92
1457-acnA w/2%glucose
1457-acnA w/4%EtOH
100
1457-acnA in DTSB
1457

Figure 7.7. a) 2D PCA scores plot and **b)** 2D OPLS-DA comparing wild-type *S. epidermidis* 1457 cells grown 6 h in standard TSB media (●), 6 h growth of aconitase mutant strain 1457-*acnA*::*tet*M in standard TSB media (▲), aconitase mutant strain 1457-*acnA*::*tet*M in iron-depleted media (DTSB) (▲), with the addition of 4% ethanol (▲), with the addition of 2% glucose (▲), with the addition of 0.06 µg/mL tetracycline (▲), and with the addition of 5% NaCl (▲). The ellipses correspond to the 95% confidence limits from a normal distribution for each cluster. The four ellipses correspond to clusters formed by (i) wild-type *S. epidermidis* 1457 cells, (ii) aconitase mutant strain 1457-*acnA*::*tet*M in standard TSB media, (iii) aconitase mutant strain 1457-*acnA*::*tet*M in DTSB or with the addition of 4% ethanol, and (iv), aconitase mutant strain 1457-*acnA*::*tet*M with the addition of 2% glucose, 0.06 µg/mL tetracycline, or 5% NaCl. For the OPLS-DA scores plot, the 6 h growth of wild-type *S. epidermidis* 1457 (●) was designated the control class and the remainder of the cells were designated as treated. The OPLS-DA used 1 predictive component and 2 orthogonal components to yield a R^2X of 0.488, R^2Y of 0.976 and Q^2 of 0.961. Metabolomic tree diagram generated from the **c)** 2D PCA scores plot depicted in (a) and (d) 2D OPLS-DA scores plot depicted in (b). The label colors match the symbol colors from the 2D scores plot. Each node is labeled with the boot-strap number, where a value above 50 indicates a statistically significant separation.

primary clusters. The wild-type cells were defined as the controls and, as expected, the aconitase mutant with and without the stress conditions formed a single large cluster. Consistent with the PCA analysis, the single cluster also contained the addition of 2% glucose. The OPLS-DA scores plot, along with the metabolomics tree diagram, suggests sub-clusters are present within the large aconitase mutant cluster. But, the aconitase mutant data is spread throughout this cluster, such that an ellipse that corresponds to the 95% confidence limit for the aconitase mutant data encompasses the two other apparent sub-clusters. This result indicates that within the resolution of the PCA and OPLS-DA model, no statistical difference is observed between the metabolomes of the aconitase mutant with and without the stress conditions. Critically, the S-plot obtained from the comparison between the wild-type cells and the aconitase mutant treated with the stress conditions was identical to the S-plot generated from the comparison between the wild-type cells and the untreated aconitase mutant cells. Again, this supports the conclusion that the addition of the stressors did not perturb the metabolome of the aconitase mutant cells.

7.3.4 Detailed analysis of changes to the S. epidermidis metabolome caused by environmental stress.

An overall correlation between the metabolomes of *S. epidermidis* under stress and TCA cycle inactivation provides further support for our hypothesis that environmental conditions induce biofilm formation though the regulation of the TCA cycle [23]. Specifically, the disparate signals of 2% glucose, 4% ethanol, 0.06 μg/mL tetracycline and iron-limitation are all sensed by the TCA cycle. To further support our

hypothesis, a detailed analysis of changes to the *S. epidermidis* metabolome caused by these environmental stress factors was necessary. We previously reported an analysis of changes in the metabolome of *S. epidermidis* caused by TCA cycle inactivation that resulted in an increase in PIA production [16]. Among other observed changes, amino-acids derived from TCA cycle intermediates (Asn, Asp, Gln, and Glu) exhibited a decrease in concentration. Correspondingly, an increase in concentrations was observed for the PIA biosynthetic precursors UDP-N-acetyl-glucosamine and fructose-6-phosphate. A similar approach using 2D ^{1}H-^{13}C HSQC and 2D ^{1}H-^{1}H TOCSY NMR spectra was applied to quantitate metabolite changes in the *S. epidermidis* metabolome caused by Fe-limitation and 4% ethanol (Figure 7.8) [23]. 2D NMR spectra improve metabolite identification by reducing the complexity and congestion of 1D ^{1}H NMR spectrum by spreading the information into two-dimensions. Additionally, the 2D ^{1}H-^{13}C HSQC experiment allows for monitoring the flow of carbon-13 through the metabolome from a specifically ^{13}C-labeled metabolite. Alternatively, the 2D ^{1}H-^{1}H TOCSY spectrum monitors all detectable metabolites with a bias to metabolites with the highest concentration. This may include the carbon-13 labeled metabolites observed in the 2D ^{1}H-^{13}C HSQC spectrum in addition to non-isotope labeled metabolites produced from other carbon sources. Therefore, the 2D ^{1}H-^{13}C HSQC and 2D ^{1}H-^{1}H TOCSY NMR spectra are complementary experiments for metabolomics and allow for a more complete analysis of metabolite concentration changes. Specifically, *S. epidermidis* wild-type cells and the aconitase mutant cells were grown with and without stress factors and harvested during either the exponential or post-exponential phase with and without the addition of ^{13}C-glucose. A total of 12 different bacterial culture conditions were prepared in triplicate

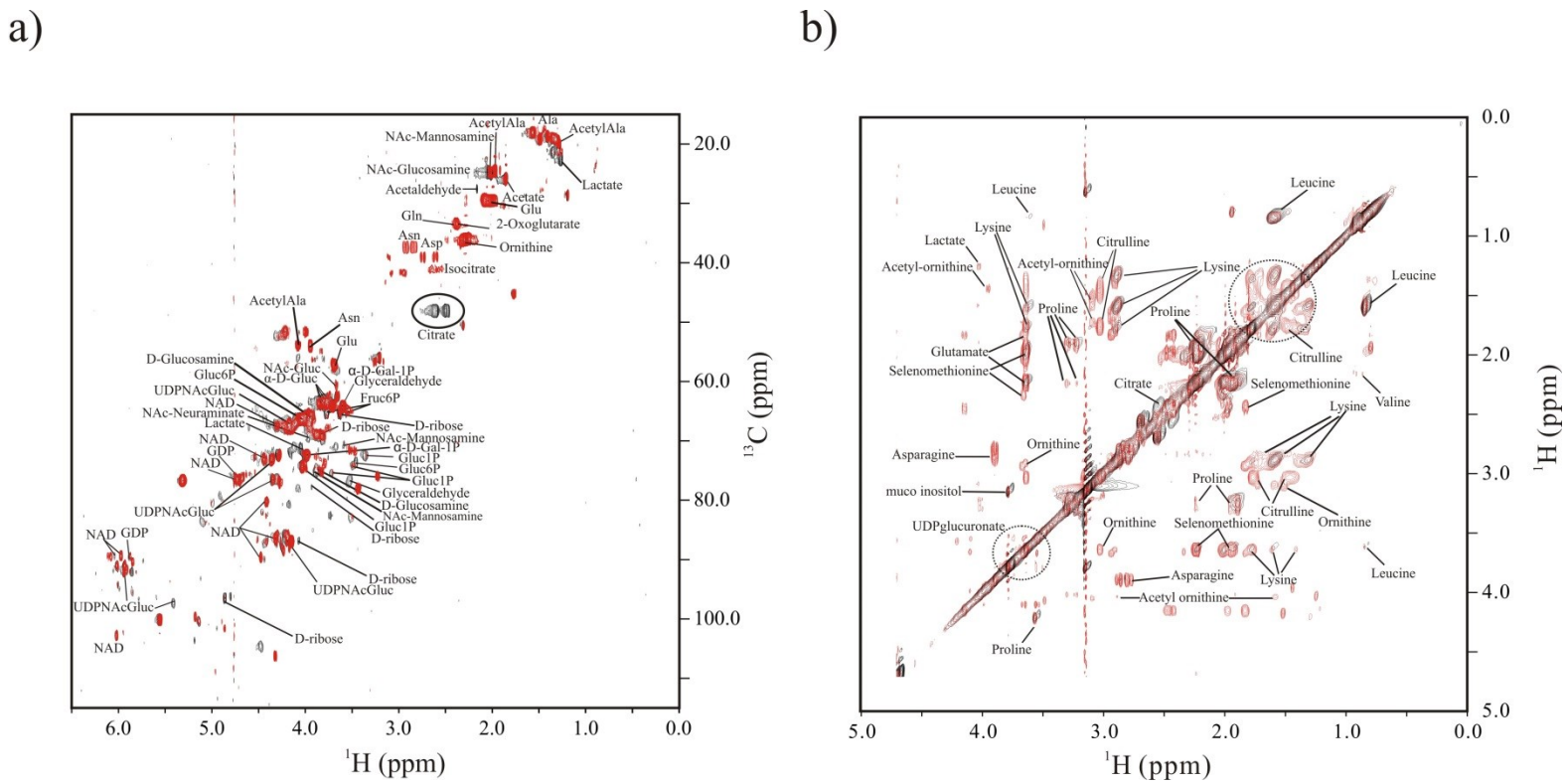


Figure 7.8. a) Overlay of 2D ^{1}H-^{13}C HSQC spectra comparing wild-type *S. epidermidis* strain 1457 (red) and aconitase mutant strain 1457-*acnA*::*tet*M (black) grown for 6 h in standard TSB media augmented with 0.25% ^{13}C-glucose. NMR resonances corresponding to specific metabolites are labeled, where citrate is circled. (b) Overlay of 2D ^{1}H-^{1}H TOCSY spectra comparing wild-type *S. epidermidis* strain 1457 (red) and aconitase mutant strain 1457-*acnA*::*tet*M (black) grown for 6 h in standard TSB media.

for both the 2D ^{1}H-^{13}C HSQC and 2D ^{1}H-^{1}H TOCSY NMR experiments for a minimum of 72 bacterial cultures or NMR experiments. To maintain consistency, *S. epidermidis* wild-type cells were used as a reference and prepared with each bacterial culture set.

Differences in metabolite concentrations between the stresses, the bacterial growth phases, the aconitase mutant cells, and the wild-type cells was based strictly on detecting changes in peak intensities in the 2D ^{1}H-^{13}C HSQC and 2D ^{1}H-^{1}H TOCSY NMR experiments. To minimize contributions from experimental variability, three levels of normalization were used. First, metabolite concentrations were normalized based on the total number of cells grown for each culture. Second, the peak intensities in each NMR spectrum were normalized by the spectrum's average peak intensity. Third, each individual peak was normalized by scaling by the largest intensity observed for that peak across the set of NMR spectra. The intensity of peaks assigned to each metabolite within a spectrum were averaged and then averaged across the triplicate NMR data set. Relative changes in peak intensities (metabolite concentrations) were compared to the *S. epidermidis* wild type metabolome and are displayed as bar graphs in Figure 7.9. Importantly, the metabolites identified from the 2D NMR experiments were also consistent with the metabolites identified as the major contributors to class distinction in the 2D OPLS-DA plots (Figures 7.5, 7.7). The OPLS-DA S-plots (Figure 7.10) identifies the relative contribution of each bin (^{1}H NMR chemical shift) to the clustering in the corresponding 2D scores plot. Each NMR bin with a high reliability (p(corr)[1] ~ 1 or -1) and a high magnitude (p[1] > 0.1 or < -0.1) was assigned to a metabolite that was also found to be present in the Figure 7.9 bar graph.

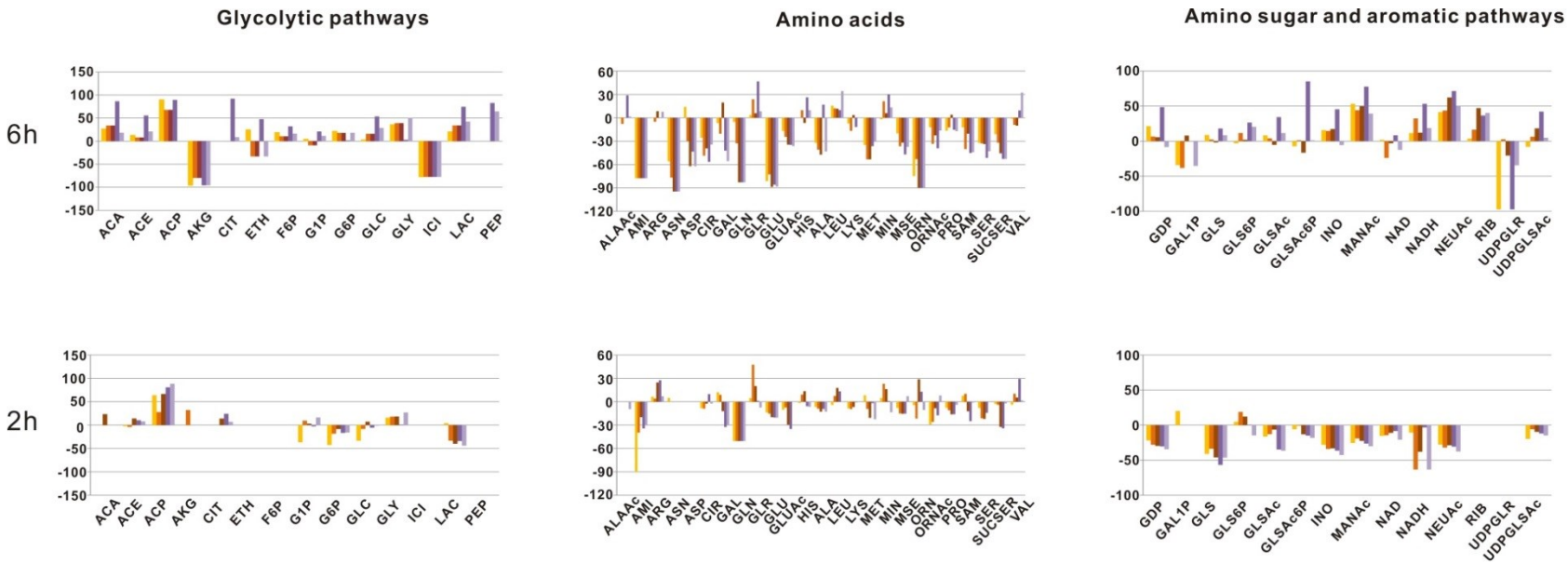

Glycolytic pathways
Amino acids
Amino sugar and aromatic pathways
6h
2h
ACA ACE ACP AKG CIT ETH F6P G1P G6P GLC GLY ICI LAC PEP
ALAAc AMI ARG ASN ASP CIR GAL GLN GLR GLU GLUAc HIS ALA LEU LYS MET MIN MSE ORN ORNAc PRO SAM SER SUCSER VAL
GDP GAL1P GLS GLS6P GLSAc GLSAc6P INO MANAc NAD NADH NEUAc RIB UDPGLR UDPGLSAc

Figure 7.9. Bar graphs showing the percent change in metabolite concentrations relative to wild-type *S. epidermidis* strain 1457 grown in standard TSB media. Metabolite concentration changes were measured after 2 h and 6 h bacterial growths for the aconitase mutant strain 1457-*acnA*::*tet*M in TSB media (■), aconitase mutant strain 1457-*acnA*::*tet*M with iron-depletion (■), aconitase mutant strain 1457-*acnA*::*tet*M with the addition of 4% ethanol (■), wild-type *S. epidermidis* strain 1457 with iron-depletion (■), and wild-type *S. epidermidis* strain 1457 with the addition of 4% ethanol (■) Positive values represent increased concentrations while negative values represent decreased concentrations with respect to *S. epidermidis* strain 1457 grown in standard TSB media. The metabolite names were abbreviated as follows: ACA (Acetaldehyde), ACE (Acetate), ACP (Acetyl-P), AKG (α-ketoglutarate), ALAAc (Acetyl-alanine), AMI (4-Aminobutanoate), ARG (Arginine), ASN (Asparagine), ASP (Aspartate), CIR (Citrulline), CIT (Citrate), ETH (Ethanol), F6P (Fruc-6P), G1P (Gluc-1P), G6P (Gluc-6P), GAL (Galacturonic-acid), GAL1P (α-D-Gala-1P), GLN (Glutamine), GLR (Glucuronate), GLS (D-glucosamine), GLS6P (Glucosamine-6P), GLSAc (N-Ac-D-glucosamine), GLSAc6P (Acetyl-glucosamine-6P), GLU (Glutamate), GLUAc (Acetyl-glutamate), GLY (Glyceraldehyde), HIS (Histidine), ICI (Isocitrate), INO (Ino, Ade, Xan), LAC (Lactate), ALA (alanine), LEU (Leucine), LYS (Lysine), MANAc (N-acetyl-D-mannosamine), MET (Methionine), MIN (myo inositol), MSE (selenomethionine), NEUAc (N-Ac-neuraminate), ORN (Ornithine), ORNAc (Acetyl-ornithine), PEP (Phosphoenolpyruvic acid), PRO (Proline), RIB (D-ribose), SAM(S-adenosyl-methionine), SER (Homoserine), SUCSER (O-Succinyl-L-homoserine), UDPGLR (UDP-glucoronate), UDPGLSAc (UDP-NAc-D-glucosamine), VAL (Valine)

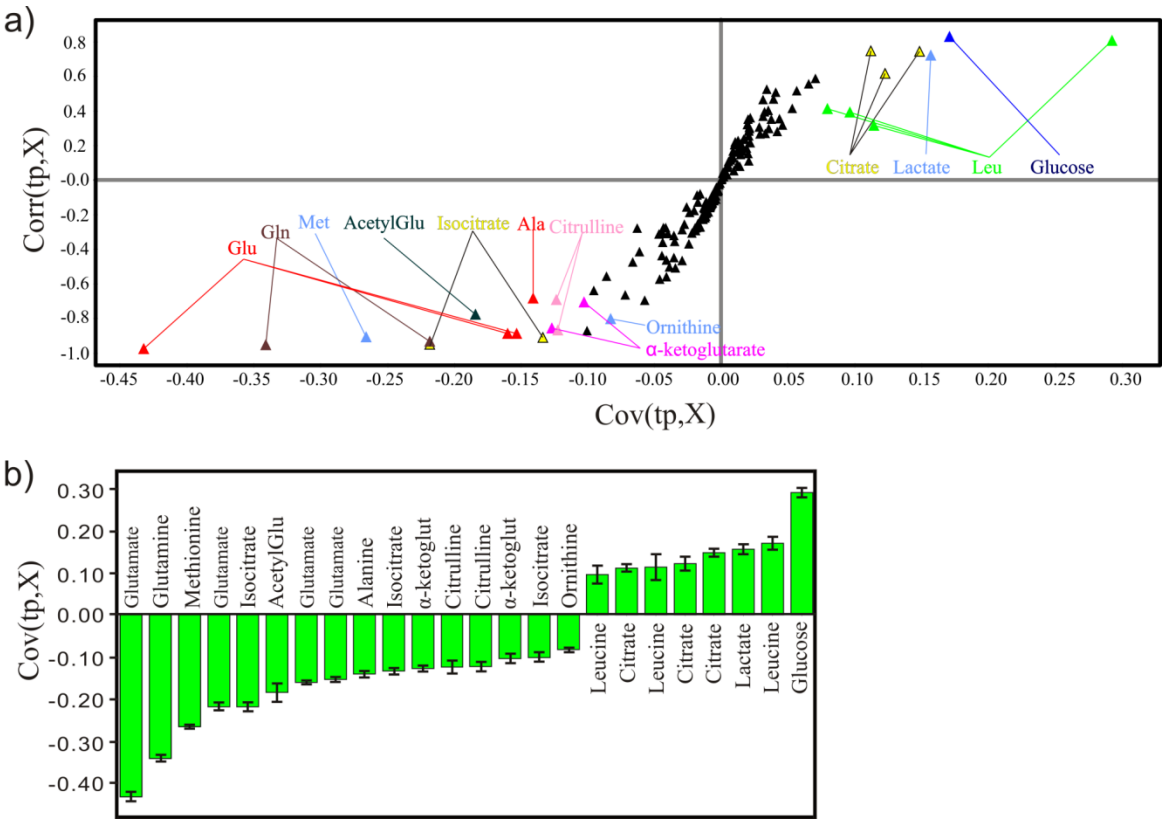


Figure 7.10. a) OPLS-DA S-plots comparing the *S.epidermidis* 1457 and the aconitase mutant 1457-*acnA*::*tet*M under the different stress conditions described in Figure 4. **b)** OPLS-DA loading plot showing the contribution of the identified metabolites from the S-plot. These results are comparable to the bar graphs depicted in figure 7.9 from the analysis of 2D NMR data.

An inactivated TCA cycle caused concentration changes for 55 metabolites involved in the amino sugar pathway, glycolytic pathway, several amino acid pathways, and the TCA cycle. The NMR data shows the amount of cellular glucose in the 6 hour post-exponential wild-type strain was reduced by 80% compared to the 2 hour exponential phase. The amount of cellular acetate was increased by 25%. As expected, the inactivation of the TCA cycle in the aconitase mutant resulted in the accumulation of a large concentration of acetate. Acetate was the most intense peak in the 6 hour cultures for the wild-type and aconitase mutant strains shown in Figure 7.8. There were also noticeable differences in the aconitase mutant metabolome. Peaks corresponding to the amino acids derived from the TCA cycle intermediates such as Asn, Asp, Gln, and Glu were not present. Not surprisingly, a large amount of citrate was also seen, since the inactivated aconitase prevents the conversion of citrate to isocitrate. Other metabolites associated with the glycolytic pathway were up-regulated. Similarly, some amino sugar and aromatic metabolites were up-regulated except for the significant down-regulation of UDP-glucuronate.

As expected from the clustering pattern in the 2D PCA scores plot, the direction of carbon flow in cells under stress were similar to the aconitase mutant cells, but dramatically different from wild-type cells during post exponential growth. A decrease in the concentration of amino acids derived from TCA cycle intermediates such as Asp, Asn, Glu, and Gln shows that the TCA cycle is still repressed when the cells are under stress. Instead the carbon flow is redirected back into the glycolytic pathway as indicated by an increase in concentrations for phosphoenolpyruvate (PEP), acetaldehyde, and fructose 6-phosphate. The carbon flow was also directed into the amino sugar pathway

with an increase in concentrations for UDP-N-acetylglucosamine, N-acetyl-neuraminate, and N-acetyl-D-mannosamine. UDP-N-acetylglucosamine is an important precursor to PIA formation. Again, the detailed analysis of changes in the metabolome of *S. epidermidis* provides additional support for the role of TCA cycle activity in a metabolic signaling pathway that transduces disparate external stimuli into internal metabolic signals that facilitate biofilm formation. Effectively, the observed changes in the metabolome caused by disparate external stimuli are consistently suppressing TCA cycle activity and inducing PIA synthesis required for biofilms.

Conversely, the change of the carbon flow in the 2 hour growth is minimal. Again, this is consistent with the low TCA activity and the similar clustering in the 2D PCA scores plot between the aconitase mutant and wild-type 2 hour growth (Figure 7.2b). Obviously, stress factors cannot suppress an already inactive TCA cycle. Instead, the catabolic conversion of glucose into intermediates throughout the glycolytic pathway proceeds as expected with a slight change since pyruvate is also produced. Much of the carbon-13 from glucose was still directed to the production of acetate. The NMR data indicates cellular acetate concentrations were similar across the 2 hour bacterial cultures, but an increased amount of acetyl-phosphate was accumulated under stress conditions. This confirms that when an abundant amount of glucose is present, it is processed by glycolysis and pyruvate dehydrogenase into acetyl-CoA and converted into acetyl-phosphate for use in substrate level phosphorylation. The excretion of acetate into the culture medium helps pH homeostasis due to the large flux of acetate [71]. Aside from the glycolytic pathway, the NMR data indicates that wild-type bacteria tend to utilize

glucose more efficiency based on small decreases in amino sugar and aromatic metabolites.

7.3.5 Metabolic rearrangements during TCA cycle stress.

The metabolome of *S. epidermidis* is not a series of independent isolated metabolic pathways, but instead is a complex inter-connected network. Thus, metabolic pathways connected to the TCA cycle are also affected by changes in TCA cycle activity. In order to visualize the cascade effect of inactivating the TCA cycle, a metabolic network was constructed using Cytoscape [40]. The metabolic network (Figure 7.11) was generated by manually associating each metabolite to its corresponding pathway from the KEGG [53] database and then using the automated biological network modules integrated into Cytoscape [72, 73]. The network connects the 37 metabolites identified by NMR whose concentrations are either increased (red) or decreased (green) by an inactivated TCA cycle. Only metabolites affected by a minimally active TCA cycle under all circumstances (aconitase mutant and stress factors) are highlighted on the network map. It is important to note that NMR is not able to identify every metabolite affected by perturbing the metabolome. The concentrations or stabilities of some metabolites are simply below the NMR detection limit. These intervening and undetected metabolites are colored grey in the metabolic network. The network shows the TCA cycle as the central pathway where common metabolites connect the urea cycle, alanine metabolism, and glycolysis/gluconeogenesis that then leads to amino sugar metabolism and other metabolites associated with PIA synthesis.

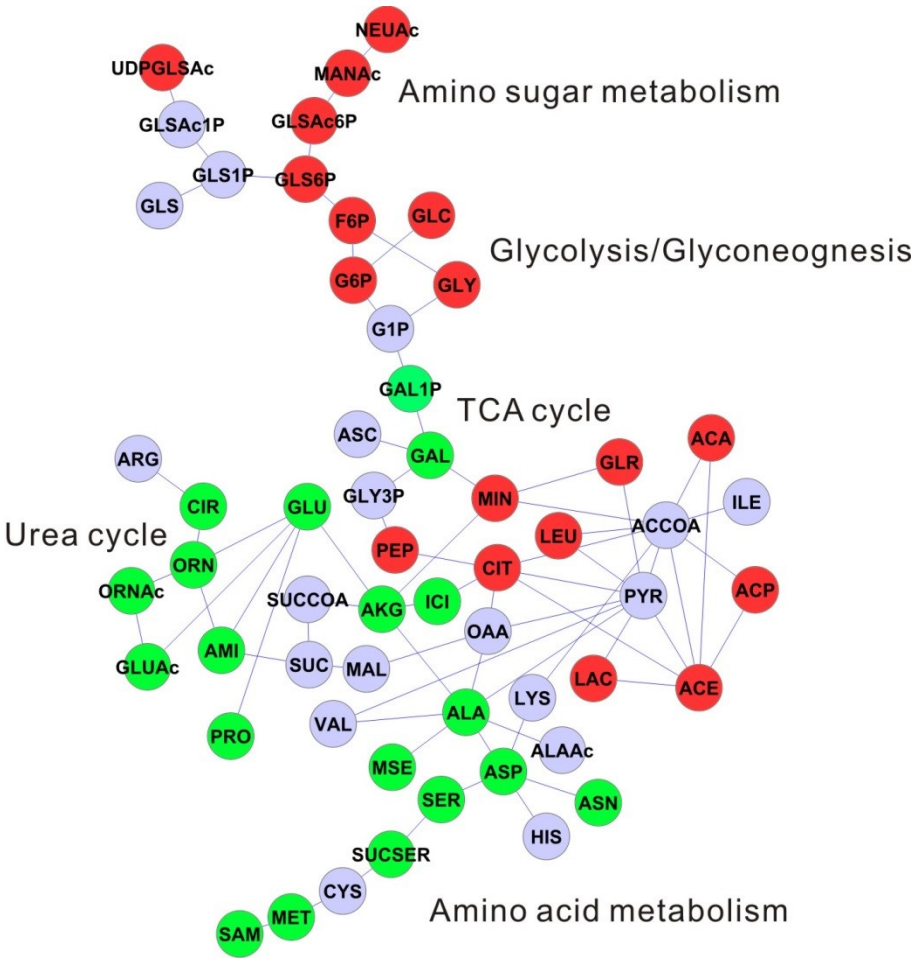


Figure 7.11. Cytoscape [40] network depicting the metabolite concentration changes caused by the inactivation of the TCA cycle. Nodes colored red correspond to metabolites with an increase in concentration due to TCA inactivated. Nodes colored green correspond to metabolites with a decrease in concentration due to TCA inactivated. Nodes colored grey correspond to metabolites that are not observed in the NMR spectra, do not have a reference NMR spectrum (or assignment), or did not exhibit a significant concentration change. Metabolic pathways are labeled on the network. The metabolite names were abbreviated as described in the legend to Figure 7.9.

7.4 Conclusion

The systematic analysis of the *S. epidermidis* metabolome using NMR provides further evidence for a metabolic signaling network for biofilm formation that involves the TCA cycle. Inactivation of the TCA cycle enables metabolic precursors to flow into pathways associated with PIA synthesis, an important component of *S. epidermidis* biofilm formation. Disparate environmental stress conditions known to induce biofilm formation were shown to perturb the metabolome of *S. epidermidis* in a manner similar to an aconitase mutant. Effectively, iron-depletion, and the addition of ethanol, tetracycline, and glucose resulted in the inactivation of the TCA cycle. Furthermore, a detailed analysis of the specific changes to the *S. epidermidis* metabolome indicates that essentially the same set of metabolites affected by TCA cycle inactivation are also affected by environmental stress conditions. A network map identified the TCA cycle as playing a central role in the proposed signaling pathway that also involves the urea cycle, alanine metabolism, glycolysis/gluconeogenesis, amino sugar metabolism and other metabolites associated with PIA synthesis. Interestingly, the addition of NaCl or autoinducer-2 did not induce any effect on the *S. epidermidis* metabolome or effect TCA cycle activity. Suggesting these factors must act through a distinct process from the other environmental factors.

7.6 References

1. Brent, R., *Cell Signaling: What is the Signal and What Information Does it Carry?* FEBS Letters, 2009. **583**(24): p. 4019-4024.

2. Jorgensen, C. and R. Linding, *Simplistic Pathways or Complex Networks?* Current Opinion Genetic Development. **20**(1): p. 15-22.

3. Kholodenko, B.N., *Spatially Distributed Cell Signaling.* FEBS Letters, 2009. **583**(24): p. 4006-4012.

4. Scott, J.D. and T. Pawson, *Cell Signaling in Space and Time: Where Proteins Come Together and When They're Apart.* Science, 2009. **326**(5957): p. 1220-1224.

5. Mascher, T., J.D. Helmann, and G. Unden, *Stimulus Perception in Bacterial Signal-Transducing Histidine Kinases.* Microbiology Molecular Biology Review, 2006. **70**(4): p. 910-938.

6. Gill, S.R., et al., *Insights on Evolution of Virulence and Resistance from the Complete Genome Analysis of an Early Methicillin-Resistant Staphylococcus aureus Strain and a Biofilm-Producing Methicillin-Resistant Streptococcus epidermidis Strain.* Journal Bacteriology, 2005. **187**(7): p. 2426-2438.

7. Novick, R.P. and E. Geisinger, *Quorum Sensing in Staphylococci.* Annual Review Genetics, 2008. **42**(1): p. 541-564.

8. Reznikoff, W.S., et al., *The Regulation of Transcription Initiation in Bacteria.* Annual Review Genetics., 1985. **19**: p. 355-87.

9. Wu, S., L.H. de, and A. Tomasz, *Sigma-B, A Putative Operon Encoding Alternate Sigma Factor of Staphylococcus aureus RNA Polymerase: Molecular Cloning and DNA Sequencing.* Journal of Bacteriology, 1996. **178**(20): p. 6036-6042.

10. Senn, M.M., et al., *Molecular Analysis and Organization of the σB Operon in Staphylococcus aureus.* Journal of Bacteriology, 2005. **187**(23): p. 8006-8019.

11. Somerville, G.A. and R.A. Proctor, *At the Crossroads of Bacterial Metabolism and Virulence Factor Synthesis in Staphylococci.* Microbiology Molecular Biology Review, 2009. **73**(2): p. 233-248.

12. Lloyd, D., *The Tricarboxylic Acid Cycle.* Process Biochemistry., 1966. **1**(9): p. 465-9.

13. Bae, T., et al., *Staphylococcus aureus Virulence Genes Identified by Bursa aurealis Mutagenesis and Nematode Killing.* Proceeding to the National Academy of Science U S A, 2004. **101**(33): p. 12312-7.

14. Begun, J., et al., *Staphylococcus aureus Virulence Factors Identified by Using a High-Throughput Caenorhabditis elegans-Killing Model.* Infections and Immunity, 2005. **73**(2): p. 872-7.

15. Coulter, S.N., et al., *Staphylococcus aureus Genetic Loci Impacting Growth and Survival in Multiple Infection Environments.* Molecular Microbiology, 1998. **30**(2): p. 393-404.

16. Sadykov, M.R., et al., *Tricarboxylic Acid Cycle-Dependent Regulation of Staphylococcus epidermidis Polysaccharide Intercellular Adhesin Synthesis.* Journal of Bacteriology, 2008. **190**(23): p. 7621-32.

17. Cramton, S.E., et al., *The Intercellular Adhesion (ica) Locus is Present in Staphylococcus aureus and is Required for Biofilm Formation.* Infections and Immunity, 1999. **67**(10): p. 5427-33.

18. Maira-Litran, T., et al., *Immunochemical Properties of the Staphylococcal Poly-N-Acetylglucosamine Surface Polysaccharide.* Infections and Immunity, 2002. **70**(8): p. 4433-40.

19. Rupp, M.E., et al., *Characterization of the Importance of Polysaccharide Intercellular Adhesin/Hemagglutinin of Staphylococcus epidermidis in the Pathogenesis of Biomaterial-Based Infection in a Mouse Foreign Body Infection Model.* Infections and Immunity, 1999. **67**(5): p. 2627-32.

20. Rupp, M.E., et al., *Characterization of Staphylococcus epidermidis Polysaccharide Intercellular Adhesin/Hemagglutinin in the Pathogenesis of Intravascular Catheter-Associated Infection in a Rat Model.* Infections and Immunity, 1999. **67**(5): p. 2656-9.

21. Fitzpatrick, F., H. Humphreys, and J.P. O'Gara, *The Genetics of staphylococcal Biofilm Formation--Will a Greater Understanding of Pathogenesis Lead to Better Management of Device-Related Infection?* Clinical Microbiology Infections, 2005. **11**(12): p. 967-73.

22. Vuong, C., et al., *Staphylococcus epidermidis Polysaccharide Intercellular Adhesin Production Significantly Increases During Tricarboxylic Acid Cycle Stress.* Journal of Bacteriology, 2005. **187**(9): p. 2967-73.

23. Sadykov, M.R., et al., *Using NMR Metabolomics to Investigate Tricarboxylic Acid Cycle Dependent Signal Transduction in Staphylococcus epidermidis.* Journal Biological Chemistry, 2010. **285**(47): p. 36616-24.

24. Knobloch, J.K.M., et al., *Biofilm Formation by Staphylococcus epidermidis Depends on Functional RsbU, An Activator of the sigB Operon: Differential Activation Mechanisms Due to Ethanol and Salt Stress.* Journal of Bacteriology, 2001. **183**(8): p. 2624-2633.

25. Campbell, I.M., et al., *In vitro Response of Staphylococcus aureus from Cystic Fibrosis Patients to Combinations of Linoleic and Oleic Acids Added to Nutrient Medium.* Journal of Clinical Microbiology, 1983. **18**(2): p. 408-15.

26. Mack, D., N. Siemssen, and R. Laufs, *Parallel Induction by Glucose of Adherence and a Polysaccharide Antigen Specific for Plastic-Adherent Staphylococcus epidermidis: Evidence for Functional Relation to Intercellular Adhesion.* Infection and Immunity, 1992. **60**(5): p. 2048-57.

27. Gerke, C., et al., *Characterization of the N-Acetylglucosaminyltransferase Activity Involved in the Biosynthesis of the Staphylococcus epidermidis Polysaccharide Intercellular Adhesin.* Journal of Biological Chemistry, 1998. **273**(29): p. 18586-18593.

28. Rachid, S., et al., *Alternative Transcription Factor Sigma(B) is Involved in Regulation of Biofilm Expression in a Staphylococcus aureus Mucosal Isolate.* Journal of Bacteriology, 2000. **182**(23): p. 6824-6.

29. Cramton, S.E., et al., *Anaerobic Conditions Induce Expression of Polysaccharide Intercellular Adhesin in Staphylococcus aureus and Staphylococcus epidermidis.* Infections and Immunity, 2001. **69**(6): p. 4079-85.

30. Deighton, M. and R. Borland, *Regulation of Slime Production in Staphylococcus epidermidis by Iron Limitation.* Infection and Immunity, 1993. **61**(10): p. 4473-9.

31. Elci, S., S. Atmaca, and K. Gul, *Effect of Iron Limitation on the Amount of Slime Produced by Strains of Staphylococcus epidermidis.* Cytobios, 1995. **84**(338-339): p. 141-6.

32. Evans, E., M.R.W. Brown, and P. Gilbert, *Iron Chelator, Exopolysaccharide and Protease Production in Staphylococcus epidermidis: A Comparative Study of the Effects of Specific Growth Rate in Biofilm and Planktonic Culture.* Microbiology, 1994. **140**(1): p. 153-7.

33. Rachid, S., et al., *Alternative Transcription Factor sigma(B) is Involved in Regulation of Biofilm Expression in a Staphylococcus aureus Mucosal Isolate.* Journal of Bacteriology, 2000. **182**(23): p. 6824-6.

34. Zhao, L., et al., *Staphylococcus aureus AI-2 Quorum Sensing Associates with the KdpDE Two-Component System to Regulate Capsular Polysaccharide Synthesis and Virulence.* Infections and Immunity, 2010. **78**(8): p. 3506-3515.

35. Forgue, P., et al., *NMR Metabolic Profiling of Aspergillus nidulans to Monitor Drug and Protein Activity.* Journal of Proteome Research, 2006. **5**(8): p. 1916-1923.

36. Halouska, S., et al., *Use of NMR Metabolomics To Analyze the Targets of D-Cycloserine in Mycobacteria: Role of D-Alanine Racemase.* Journal of Proteome Research, 2007. **6**(12): p. 4608-4614.

37. Werth, M.T., et al., *Analysis of Metabolomic PCA Data using Tree Diagrams.* Analytical Biochemistry, 2010. **399**(1): p. 56-63.

38. Lindon, J.C., E. Holmes, and J.K. Nicholson, *Pattern Recognition Methods and Applications in Biomedical Magnetic Resonance.* Progress in Nuclear Magnetic Resonance Spectroscopy, 2001. **39**(1): p. 40.

39. Stoyanova, R. and T.R. Brown, *NMR Spectral Quantitation by Principal Component Analysis.* NMR Biomed, 2001. **14**(4): p. 271-7.

40. Killcoyne, S., et al., *Cytoscape: A Community-Based Framework for Network Modeling.* Methods Molecular Biology, 2009. **563**: p. 219-39.

41. Cao, J.G. and E.A. Meighen, *Purification and Structural Identification of an Autoinducer for the Luminescence System of Vibrio harveyi.* Journal Biological Chemistry, 1989. **264**(36): p. 21670-6.

42. Semmelhack, M.F., et al., *An Expeditious Synthesis of DPD and Boron Binding Studies.* Organic Letters, 2005. **7**: p. 569-572.

43. Nguyen, B.D., et al., *SOGGY: Solvent-Optimized Double Gradient Spectroscopy for Water Suppression. A Comparison with Some Existing Techniques.* Journal of Magnetic Resonance, 2007. **184**(2): p. 263-74.

44. Palmer, A.G.I., et al., *Sensitivity Improvement in Proton-Detected Two-Dimensional Heteronuclear Correlation NMR spectroscopy.* Journal of Magnetic Resonance, 1991. **93**(1): p. 20.

45. Kay, L.E., P. Keifer, and T. Saarinen, *Pure Absorption Gradient Enhanced Heteronuclear Single Quantum Correlation Spectroscopy with Improved Sensitivity.* Journal American Chemical Society, 1992. **114**(26): p. 3.

46. Piotto, M., V. Saudek, and V. Sklenar, *Gradient-Tailored Excitation for Single-Quantum NMR Spectroscopy of Aqueous Solutions.* Journal of Biomolecular NMR, 1992. **2**(6): p. 661-5.

47. Halouska, S. and R. Powers, *Negative Impact of Noise on the Principal Component Analysis of NMR Data.* Journal of Magnetic Resonance, 2006. **178**(1): p. 88-95.

48. Johnson, B.A., *Using NMRView to Visualize and Analyze the NMR Spectra of Macromolecules.* Methods Molecular Biology, 2004. **278**: p. 313-52.

49. Xia, J., et al., *Metabominer-Semi-Automated Identification of Metabolites from 2D NMR Spectra of Complex Biofluids.* BMC Bioinformatics., 2008. **9**: p. No pp. given.

50. Cui, Q., et al., *Metabolite Identification via the Madison Metabolomics Consortium Database.* National Biotechnology, 2008. **26**(2): p. 162-4.

51. Ulrich, E.L., et al., *BioMagResBank.* Nucleic Acids Research, 2008. **36**(Database Iss): p. D402-D408.

52. Wishart, D.S., et al., *HMDB: A Knowledge-Base for the Human Metabolome.* Nucleic Acids Research, 2009. **37**(Database Iss): p. D603-D610.

53. Kanehisa, M., et al., *KEGG for Linking Genomes to Life and the Environment.* Nucleic Acids Research, 2008. **36**(Database Iss): p. D480-D484.

54. Caspi, R., et al., *The MetaCyc Database of Metabolic Pathways and Enzymes and the BioCyc Collection of Pathway/Genome Databases.* Nucleic Acids Research, 2010. **38**(Database Iss): p. D473-D479.

55. van den Berg, R.A., et al., *Centering, Scaling, and Transformations: Improving the Biological Information Content of Metabolomics Data.* BMC Genomics, 2006. **7**: p. 142.

56. Somerville, G.A., et al., *Staphylococcus aureus Aconitase Inactivation Unexpectedly Inhibits Post-Exponential-Phase Growth and Enhances Stationary-Phase Survival.* Infections and Immunity., 2002. **70**(11): p. 6373-6382.

57. Mack, D., et al., *Mechanisms of Biofilm Formation in Staphylococcus epidermidis and Staphylococcus aureus: Functional Molecules, Regulatory Circuits, and Adaptive Responses.* International Journal Medical Microbiology, 2004. **294**(2-3): p. 203-212.

58. Collins, F.M. and J. Lascelles, *The Effect of Growth Conditions on Oxidative and Dehydrogenase Activity in Staphylococcus aureus.* Journal General Microbiology, 1962. **29**: p. 531-5.

59. Costerton, J.W., P.S. Stewart, and E.P. Greenberg, *Bacterial Biofilms: A Common Cause of Persistent Infections.* Science, 1999. **284**(5418): p. 1318-22.

60. Kornmann, H., et al., *Influence of Residual Ethanol Concentration on the Growth of Gluconacetobacter xylinus I 2281.* Applied Microbiology Biotechnology, 2003. **62**(2-3): p. 168-73.

61. Throup, J.P., et al., *The srhSR Gene Pair from Staphylococcus aureus: Genomic and Proteomic Approaches to the Identification and Characterization of Gene Function.* Biochemistry, 2001. **40**(34): p. 10392-401.

62. Tomlins, R.I., M.D. Pierson, and Z.J. Ordal, *Effect of Thermal Injury on the TCA Cycle Enzymes of Staphylococcus aureus MF 31 and Salmonella typhimurium 7136.* Canadian Journal of Microbiology, 1971. **17**(6): p. 759-65.

63. Miller, M.B. and B.L. Bassler, *Quorum Sensing in Bacteria.* Annual. Review of Microbiology, 2001. **55**: p. 165-199.

64. Lim, Y., et al., *Control of Glucose- and NaCl-Induced Biofilm Formation by rbf in Staphylococcus aureus*. Journal of Bacteriology, 2004. **186**(3): p. 722-9.

65. Cue, D., et al., *Rbf promotes Biofilm formation by Staphylococcus aureus via Repression of icaR, A Negative Regulator of icaADBC.* Journal of Bacteriology, 2009. **191**(20): p. 6363-6373.

66. Heilmann, C., et al., *Molecular Basis of Intercellular Adhesion in the Biofilm-Forming Staphylococcus epidermidis.* Molecular Microbiology, 1996. **20**(5): p. 1083-1091.

67. Hoffman, L.R., et al., *Aminoglycoside Antibiotics Induce Bacterial Biofilm Formation.* Nature, 2005. **436**(7054): p. 1171-1175.

68. Chang, Y.-M., et al., *Structural Study of TcaR and Its Complexes with Multiple Antibiotics from Staphylococcus epidermidis.* Proceedings to the National Academy of Science, 2010. **107**(19): p. 8617-8622, S8617/1-S8617/13.

69. Sousa, C., et al., *Staphylococcus epidermidis Glucose Uptake in Biofilm Versus Planktonic Cells.* World Journal Microbiology Biotechnology, 2008. **24**(3): p. 423-426.

70. Croes, S., et al., *Staphylococcus aureus Biofilm Formation at the Physiologic Glucose Concentration Depends on the S. aureus Lineage.* BMC Microbiology, 2009. **9**: p. No pp. given.

71. Somerville, G.A., et al., *Correlation of Acetate Catabolism and Growth Yield in Staphylococcus aureus: Implications for Host-Pathogen Interactions.* Infections and Immunity, 2003. **71**(8): p. 4724-32.

72. Cline, M.S., et al., *Integration of Biological Networks and Gene Expression Data using Cytoscape.* Nature Protocol, 2007. **2**(10): p. 2366-82.

73. Shannon, P., et al., *Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks.* Genome Research, 2003. **13**(11): p. 2498-504.

CHAPTER 8

CCPA REGULATES ARGININE BIOSYNTHESIS IN *STAPHYLOCOCCUS AUREUS* THROUGH REPRESSION OF PROLINE CATABOLISM[§]

8.1 Introduction

Staphylococcus aureus is a common cause of skin and soft tissue infections; however more serious complications such as bacteremia, osteomyelitis, endocarditis, and necrotizing pneumonia can occur [1]. During infection, *S. aureus* must catabolize diverse carbon sources including carbohydrates, proteins and lipids; therefore, multiple global regulators, including CcpA and CodY, subsequently regulate carbon flow [2-4]. Thus, regulation of carbon flow through central metabolism and other metabolic pathways has a direct link to expression and synthesis of virulence factors [5-7].

It has been known for over 70 years that *S. aureus* exhibits multiple amino acid auxotrophies, including arginine, valine, proline, cysteine, and leucine [8, 9]. Complicating the picture, in 1937, Gladstone demonstrated that multiple strains of *S. aureus* could be trained to grow in a chemically-defined broth lacking all twenty amino acids through extended incubation [9]. These data suggested that *S. aureus* was indeed a prototroph but repressed biosynthesis of certain amino acids. In support of this, bioinformatics analyses of the *S. aureus* genome revealed an apparently complete repertoire of biosynthetic operons needed to synthesize all 20 amino acids [10]. Included

---

[§] Chapter 8 was adapted from Nuxoll, A., *et. al*., CcpA Regulates Arginine Biosynthesis in Staphylococcus aureus through Repression of Proline Catabolism. *PLoS Pathogens* (2012), **8** (11). Reprinted with permission, copyright 2012 PLOS.

in these are the genes encoding the arginine biosynthetic pathway *argJBCDFGH* where arginine is synthesized from glutamate [11]. This pathway is highly conserved among a wide array of bacteria, including *Escherichia coli*, *Salmonella enterica* serotype *Typhimurium*, *Proteus mirabilis*, *Bacillus subtilis*, and *Streptomyces clavuligerus* among others [12-14].

Although *B. subtilis* synthesizes proline from glutamate [11, 15, 16], *S. aureus* preferentially utilizes arginine rather than glutamate as a precursor for proline biosynthesis via arginase (RocF), ornithine aminotransferase (RocD), and P5C reductase (ProC) [17]. Furthermore, Li and colleagues recently reported that proline biosynthesis is regulated through CcpA-mediated carbon catabolite repression at both *rocF* and *rocD* [18]. Carbon catabolite repression allows bacteria to preferentially utilize preferred carbon sources and therefore increase the organism's fitness [19]. The *trans*-acting carbon catabolite protein CcpA in a complex with Hpr binds to *cis*-acting DNA sequences known as catabolite responsive elements (CRE) [20-23]. In the presence of a preferred carbon source, HprK phosphorylates the Ser-46 position of Hpr and once phosphorylated, Hpr binds to CcpA [23-25].

In this study, we utilized genetic and biochemical approaches to examine arginine auxotrophy in *S. aureus*. *bursa aurealis* transposon mutagenesis identified CcpA as a regulator of arginine biosynthesis. However, instead of de-repressing the conserved arginine biosynthesis pathway (ArgJBCDFGH) via glutamate, *S. aureus* JE2 *ccpA* synthesized arginine from proline via the urea cycle. To the best of our knowledge, this is the first report of bacteria utilizing proline for arginine biosynthesis, which may indicate

a predilection to degrade and utilize proteins rich in proline (i.e. collagen) during an *S. aureus* infection for use in arginine biosynthesis. Utilization of proline to synthesize arginine demonstrates the resourcefulness of *S. aureus* and its ability to rapidly evolve to utilize nutrients that are readily available in the environment.

8.2 Materials and Methods

8.2.1 Ethics

The clinical *S. aureus* strains used in this study originated from the University of Nebraska Medical Center. The Institutional Review Board at the University of Nebraska Medical Center is charged with reviewing all research involving human subjects. The clinical *S. aureus* strains utilized in the study were de-identified and analyzed anonymously and were therefore exempt from human research committee approval. Animal experimentation was performed under a University of Nebraska Medical Center approved Institutional Animal Care and Use Committee (IACUC) Protocol to TK. The University of Nebraska Medical Center is accredited by the Association for Assessment and Accreditation of Laboratory Animal Care International (AALAC). In addition, all animals at the University of Nebraska Medical Center are maintained in accordance with the Animal Welfare Act and the DHHS ‘‘Guide for the Care and Use of Laboratory Animals.’’

Table 8.1. Bacterial Strains and Plasmids used in study

Bacterial Strain or Plasmid	Relevant Phenotype	Source
pFA545	Encodes pBursa transposase; Tet^r	[26]
pBursa	Encodes bursa aurealis transposon; Erm^r, Cam^r	[26]
pNF266	pCN51 (60) with *ccpA* cloned into the SphI and BamHI site using primers 2250 and 2251 removing *cadA*; Amp^r, Erm^r	This study
pMRS44	Derivative of pTS1-d (58) with *ccpA::ermB* Amp^r, Cam^r, Erm^r	This study
S. aureus RN4220	Restriction deficient NCTC8325-4	[27]
S. aureus Newman	S. aureus ATCC 25904	[28]
S. aureus MST14	Newman containing an allelic replacement mutation (*tetL*) in *ccpA*. Tet^r	[2]
S. aureus LAC	USA300 PFGE type; ST8. Erm^r, Cam^r, Tet^r	[29]
S. aureus JE2 Cam^s, Tet^s	LAC cured of all 3 native plasmids; Erm^s	This study
S. aureus JE2 *ccpA::tetL*	Allelic replacement mutation in *ccpA*; transduced from MST14; Tet^r	This study
S. aureus SA564 *ccpA::ermB*	Allelic replacement mutation in *ccpA* Erm^r	This study
S. aureus JE2 *ccpA::ermB*	*ccpA* mutant; transduced from SA564 *ccpA::ermB*; Erm^r	This study
S. aureus RN4220 *ccpA::tetL*	Allelic replacement mutation in *ccpA*: transduced from MST14; Tet^r	This study
JE2 *ccpA::tetL/argJ::*φNΣ	*bursa aurealis argJ* mutation in *ccpA::tetL* background.	This study
JE2 *ccpA::tetL/argB::*φNΣ	*bursa aurealis argB* mutation in *ccpA::tetL* background.	This study
JE2 *ccpA::tetL/argC::*φNΣ	*bursa aurealis argC* mutation in *ccpA::tetL* background.	This study

Table 8.1 (continued)

Bacterial Strain or Plasmid	Relevant Phenotype	Source
JE2 *ccpA::tetL/argF::*φNΣ	*bursa aurealis argF* mutation in *ccpA::tetL* background.	This study
JE2 *ccpA::tetL/argG::*φNΣ	*bursa aurealis argG* mutation in *ccpA::tetL* background.	This study
JE2 *ccpA::tetL/argH::*φNΣ	*bursa aurealis argH* mutation in *ccpA::tetL* background.	This study
JE2 *ccpA::tetL/proC::*φNΣ	*bursa aurealis proC* mutation in *ccpA::tetL* background.	This study
JE2 *ccpA::tetL/putA::*φNΣ	*bursa aurealis putA* mutation in *ccpA::tetL* background.	This study
JE2 *ccpA::tetL/rocD::*φNΣ	*bursa aurealis rocD* mutation in *ccpA::tetL* background.	This study
JE2 *ccpA::tetL/arcB1::*φNΣ	*bursa aurealis arcB1* mutation in *ccpA::tetL* background.	This study
JE2 *ccpA::tetL/arcB2::*φNΣ	*bursa aurealis arcB2* mutation in *ccpA::tetL* background.	This study
RN4220 *argH::*φNΣ	*bursa aurealis argH* mutation	This study
RN4220 *argF::*φNΣ	*bursa aurealis argF* mutation	This study
RN4220 *ccpA::tetL/argG::*φNΣ	*bursa aurealis argG* mutation in *ccpA::tetL* background	This study
RN4220 *ccpA::tetL/argH::*φNΣ	*bursa aurealis argH* mutation in *ccpA::tetL* background	This study
Newman *ccpA::tetL/argG::*φNΣ	*bursa aurealis argG* mutation in *ccpA::tetL* background	This study
Newman *ccpA::tetL/argH::*φNΣ	*bursa aurealis argH* mutation in *ccpA::tetL* background	This study

8.2.2 Bacterial Strains and Culture Conditions

For determination of arginine auxotrophy, eighty-two *S. aureus* isolates were obtained from a previous collection testing the prevalence of heterogeneous vancomycin intermediate susceptibility [30]. Other bacterial strains used in the study are shown in Table 8.1. Defined *bursa aurealis* transposon mutants were acquired from the Nebraska transposon mutant library via the Network on Antimicrobial Resistance in Staphylococcus aureus (NARSA; http://www.narsa.net). Bacterial strains were grown in either Tryptic Soy Broth (TSB; Becton Dickinson, Franklin Lakes, NJ) or Complete Defined Medium (CDM) as previously described except containing 0.25% glucose [31]. Cultures were grown aerobically (1:10 volume to flask ratio) at 37^{o}C, 250 rpm unless otherwise stated. To train JE2 to grow on media lacking arginine, cultures were grown in CDM-R broth for 6 days, at which point the bacteria were inoculated to CDM-R agar. To study the reversion frequency of JE2, Newman and RN4220, the bacteria were grown for 20 hours in 3 mL of CDM. Cells were pelleted, resuspended in 0.9% NaCl, and diluted onto CDM, CDM-R, or CDM-P. After 72 hours the colonies were counted and reversion frequency was determined by taking the number of prototrophic revertants divided by total number of colonies plated on CDM.

To determine the growth characteristics in CDM-R containing various alternative carbon sources, JE2 was grown in 3 mL of CDM overnight, pelleted and resuspended in 0.9% NaCl. 3 mL of CDM and CDM-R supplemented with either 0.25% of glucose, fructose, glycerol, sucrose, mannitol, maltose, salicin, gluconic acid, arabinose, sorbitol, or ribose (all purchased from Sigma-Aldrich, St. Louis, MO) were inoculated in a 14 mL

polypropylene round-bottom tube (Becton Dickinson) to an OD_{600} of 0.05. Cultures were grown for 18 h at 37^{o}C to stationary phase.

8.2.3 Screening of Random bursa aurealis Transposon Mutant Library

Random *bursa aurealis* transposon mutants were generated using plasmids pBursa and pFA545 and identified using inverse PCR as previously described [26]. Mutants were grown and collected in a 96 well format and pelleted and resuspended in 50 uL of 0.9% NaCl. 2 uL were plated on CDM and CDM-R and incubated at 37^{o}C for 72 hours. Approximately 2700 mutants were screened; colonies that grew on CDM-R plates were confirmed by growing in CDM-R broth.

8.2.4 Transduction, ccpA Mutant Construction and Complementation

bursa aurealis transposon mutations were moved to other strain backgrounds through transduction using phage 80α or φ11 as previously described [32]. All primers (Table 8.2) used for construction and confirmation of the *ccpA* mutation were generated based on the sequence of *S. aureus* strain Mu50 (NC_002758.2). The *ccpA* mutant was constructed by replacing a 0.6 kb internal region of the *ccpA* gene with an erythromycin resistance cassette (*ermB*) using the gene splicing by overlap extension (gene SOEing) technique [33]. *ermB* was amplified from pEC4 [34] using primers SAV1736-ermB-f and SAV1736-ermB-r, which contain sequences homologous to the *ccpA* gene. Primers BamHI-SAV1737-f and ermBSAV1736-r were used for amplification of a 1.3 kb region upstream of the *ccpA* gene. Primers ermB-SAV1736-f and SacI-acuC-f were used to

Table 8.2. Oligonucleotides used in study

Primer Name	Sequence	Relevant Characteristic
1868	TCTACAAGAACGCGATGTGC	Forward primer used to amplify *argF*
1869	TTCACCCAATGTAGCAACCA	Reverse primer used to amplify *argF*
1872	CAAAATCTATGGGGCAGAGC	Forward primer used to amplify *argG*
1873	CCATGCAACATCGCATTTAC	Reverse primer used to amplify *argG*
1874	TGCAACTATGCTTGCGAATC	Forward primer used to amplify *argH*
1875	TGCTAGTTCCGTTGCATTTG	Reverse primer used to amplify *argH*
1876	ACCCCCTACTTCAAGGCACT	Forward primer used to amplify *argD*
1877	CGTCTTGAAAAGCTGCAACA	Reverse primer used to amplify *argD*
1878	TAGGTATCGTTGGCGGTAGC	Forward primer used to amplify *argC*
1879	CTCGATTTCCGGTTTGTGTT	Reverse primer used to amplify *argC*
1880	ACACAAACATGGGTTGCTCA	Forward primer used to amplify *argJ*
1881	GTTTCCCATTGTGGATGGTC	Reverse primer used to amplify *argJ*
1882	AACACACGCTCATTGCAGAC	Forward primer used to amplify *argB*
1883	AGGACAGCCATTTTCAATCG	Reverse primer used to amplify *argB*
1929	TTGCAGCGCATGATCAAGGT	Forward qRT-PCR; *argF*
1930	TTCCACACTGGTACGCCTGAA	Reverse qRT-PCR; *argF*

Table 8.2 (continued)

Primer Name	Sequence	Relevant Characteristic
1933	TGGGCATGGAGTCGTGAAGA	Forward qRT-PCR; *argG*
1934	CTCTGGTGGCGCAGCATAAG	Reverse qRT-PCR: *argG*
1935	ATCGAAGGCAGCATTGCACA	Forward qRT-PCR; *argH*
1936	CCACCAGCATCACCGATACG	Reverse qRT-PCR; *argH*
1937	TGGGAGCAAGTCGTTCCAGA	Forward qRT-PCR; *argD*
1938	CAAGCGCTGCCGTCGATATT	Reverse qRT-PCR; *argD*
1939	TCGTTGGCGGTAGCGGTTAT	Forward qRT-PCR; *argC*
1940	GGTGCTGGTGTCGCAAAGA	Reverse qRT-PCR; *argC*
1941	CATCGTCTTTGGCAAGTGCAG	Forward qRT-PCR; *argJ*
1942	CAGCTGTGCATCCAAACCACA	Reverse qRT-PCR; *argJ*
1943	TGGCGGTGGCCCATTTATTA	Forward qRT-PCR; *argB*
1944	CCCCAGTTGAAGCAACAGCA	Reverse qRT-PCR; *argB*
2250	CCCGGG*GCATGC*TGAATTCCA AGCATTTTATGATGA	Forward *ccpA;* contains SphI* restriction site
2251	CCCGGG*GGATCC*TGAATTTATT TTGTAGTTCCTCGGTA	Reverse *ccpA;* contains BamH1* restriction site
2186	TCGCTAAATTTTTCCAAACAAAA	Forward *ccpA*
2187	AAAGGCATTCTTCCAACACC	Reverse *ccpA*

Table 8.2. (continued)

Primer Name	Sequence	Relevant Characteristic
1888	TTTGGGCATATGTTAACGACAG AAAAACTAGTT	Forward *hprK*
1889	GGGGGATCCCTACTCCTCACTC TTATGACTG	Reverse *hprK*
1886	AATGTACATATGGAACAAAATTC ATATGTAATC	Forward *ptsH*
1887	ACAGGATCCTTTAGTCAATCCTT CTTTTGATAAGAC	Reverse *ptsH*
BamHI-SAV1737F	CCAGGATCCACGTATCACGTA AGTTGAAACCTGAAG	*ccpA* SOEing
ermB-SAV1736R	CAACATGACGAATCCCTCCTTC GCGCTTCTCTTGCTACATCATAT ATAGTAAC	*ccpA* SOEing
SAV1736-ermBF	CTATATATGATGTAGCAAGAG AAGCGCGAAGGAGGGATTCG TCATGTTGG	*ccpA* SOEing
SAV1736-ermBR	CAATTCAATGTATCACCTAAT CAATTCAATGTATCACCTAAT ATTATTTCCTCCCG	*ccpA* SOEing
ermB-SAV1736F	CGGGAGGAAATAATTCTATGA GTCGCGGCCTTCAATTAGGTG ATACATTGAATTG	*ccpA* SOEing
SacI-AcuC-F	GCAGAGCTCAGCAACAAGCGT TTGATGATATTCG	*ccpA* SOEing
SAV1737F	GCAACAAAGGACCATTTAACGA TAATAC	*ccpA* mutant confirmation
AcuC F	GGTGGACTTGAAATATTCGCTA CAG	*ccpA* mutant confirmation
ermB F	GGGTCAATCGAGAATATCGTCA ACTG	*ccpA* mutant confirmation
ermB R	GCCCTTTACCTGTTCCAATTTCG	*ccpA* mutant confirmation

*Restriction denotes in italics

amplify a 1.7 kb region downstream of the *ccpA* gene. The resulting 4.1-kb PCR product contained BamHI and SacI sites that were used for cloning into pTS1-d [35] to generate plasmid pMRS44. Plasmid pMRS44 was used to construct *S. aureus* SA564 *ccpA::ermB* using the temperature shift protocol as previously described [36]. Allelic replacement of the internal region of the *ccpA* gene by the *ermB* cassette was verified by PCR using primers ermB-f, ermB-r, SAV1737-f and acuC-f. The *ccpA::ermB* mutation was subsequently moved to JE2 through phage 80α transduction and confirmed using primers noted above. For the *ccpA* complementation plasmid pNF266, *ccpA* was amplified from JE2 using primers 2250 and 2251 (Table 8.2), digested with SphI and BamHI, and cloned into pCN51 [37]. Note that two *ccpA* mutants were constructed in this study, JE2 *ccpA::ermB* and JE2 *ccpA::tetL*. JE2 *ccpA::tetL* was generated by phage 80α transduction of the *ccpA::tetL* allele from MST14 [2] so double mutants could be constructed using *ermB* as the second selectable marker.

8.2.5 NMR Data Collection

JE2 and JE2 *ccpA::ermB* were grown in 50 mL CDM to stationary phase. JE2 and JE2 *ccpA::ermB* were subsequently inoculated to an OD_{600} of 0.05 in CDM containing 100 mM of either $^{13}C_5$-glutamate or $^{13}C_5$-proline (Isotec) and grown to stationary phase. Cultures were normalized to an OD_{600} of 2.0 and pelleted by centrifugation (3000 rpm, 20 minutes, 4^{o}C). Pellets were subsequently washed in 10 mL of cold sterile water and resuspended in 1 mL cold sterile water. The pellet was lysed using a bead beater (MP Biomedicals) and centrifuged for 15 minutes at 13,000 rpm at 4^{o}C. This lysis step was

repeated two more times and the pellet frozen in an ethanol/dry ice bath. The samples were then lyophilized, suspended in 600 μL of 50 mM phosphate buffer (pH=7.2, uncorrected) in 99.8% D2O (Isotec), and transferred to 5 mm NMR tubes for analysis. The NMR spectra were collected on a Bruker 500 MHz Avance spectrometer equipped with a triple resonance, Z-axis gradient probe. A BACS-120 sample changer with Bruker Icon software was used to automate the NMR data collection. The 2D ^{1}H-^{13}C HSQC spectra were collected with a standard Bruker pulse sequence (HSQCETGP), solvent presaturation and a relaxation delay of 1.5s. Each 2D ^{1}H-^{13}C HSQC spectrum was collected with a spectrum width of 4734.85 Hz and 2048 data points in the direct (^{1}H) dimension; and a spectrum width of 13834.26 Hz and 64 data points in the indirect (^{13}C) dimension. A total of 16 dummy scans and 128 scans were used to obtain each 2D ^{1}H-^{13}C HSQC spectra.

The spectra were processed using the NMRPipe software package [38]. The spectra were Fourier transformed, manually phased, and baseline corrected. The processed 2D ^{1}H-^{13}C HSQC spectra were then analyzed using NMRView [39] to assign chemical shifts and intensities to each peak. The chemical shift list were assigned to specific metabolites using the Human Metabalome Database [40], Madison Metabolomics Database [41], and Platform for Riken Metabolomics [42] with a tolerance level of 0.05 ppm and 0.40 ppm in the ^{1}H and ^{13}C chemical shifts respectively. The presence of metabolites and metabolomics pathways was verified using the Kyoto Encyclopedia of Genes and Genomes (KEGG) [43] and MetaCyc [44] databases. The quantification of metabolomic peak intensities were performed in a similar manner as

previously described [45]. The relative percent concentration difference was determined by subtracting averages from the two cultures. A student T-test was performed to verify the significance at a 95% confidence level, of the relative percent concentration differences.

8.2.6 RNA Isolation and Northern Blot Analysis

Cultures of *S. aureus* JE2 and JE2 *ccpA::ermB* were grown overnight in CDM, diluted to an OD_{600} of 0.05 into fresh CDM or CDM-R (1:10 volume to flask ratio, 250 rpm), and grown at 37^{o}C to an OD_{600} of 1.5 (mid-exponential growth). Cells were pelleted at 3000 x g for 20 minutes at 4^{o}C and resuspended in RLT buffer with 1% b-mercaptoethanol. Next, they were transferred to lysing matrix B tubes (MP Biomedicals) and processed in a FP120 FastPrep cell disrupter (MP Biomedicals) for 24 seconds at a setting of 6.0. The cells were pelleted at 13000 rpm at 4^{o}C for 15 minutes; top-phase was combined with 500 µL of ethanol. The samples were then processed using an RNeasy mini kit, according to manufactures instructions (Qiagen, Inc.). Primers listed in Table 8.2 were used to make DNA probes that were subsequently labeled with digoxigenin-labeled dUTP (Roche). 5 µg of RNA was used for northern analysis that was performed using DIG buffers and washes (Roche). Anti-Digoxigenin-AP Fab fragments (Roche) was used with ECF substrate (GE Healthcare) for detection. Blots were visualized using the Typhoon FLA 7000 imaging system (GE Healthcare).

8.2.7 Amino Acid Analysis

JE2 and JE2 *ccpA::ermB* were grown overnight in 50 mL (500 mL flask) of CDM. Cultures were inoculated to a starting OD_{600} of 0.05 in CDM (100 mL in 1 L flask, 250 rpm, 37^{o}C) and grown for 5 hours. 500 µL of media was collected and pelleted for 5 minutes at maximum speed. Supernatant was collected and filtered through 3,000 MWCO Amicon Ultra centrifugal filters (Millipore) according to manufactures instructions. Amino acid analysis was performed by the Protein Structure Core Facility, UNMC, using a Hitachi L-8800.

8.2.8 Animal Models

Mouse subcutaneous abscess model. Subcutaneous abscesses were established in C57BL/6 mice following the injection of 5 x 10^5 CFU of *S. aureus* JE2. Tissues were collected at day 7 post infection and processed for immunofluorescence staining for either type I collagen (Millipore, Billerica, MA), inducible nitric oxide synthase (Abcam, Cambridge, MA), or arginase (Santa Cruz, San Diego, CA.) For type I collagen, tissues were incubated with the nuclear stain DAPI to accentuate the abscess core. *Mouse kidney abscess model.* C57BL/6 mice were anesthetized using ketamine and xylazine and 100 µL containing 10^7 CFU of *S. aureus* JE2, JE2 *argH*::φNΣ, or JE2 *argF*::φNΣ were inoculated retro-orbitally. On day 20 following inoculation, the animals were sacrificed and the kidneys were excised, homogenized, and subsequently plated for bacteriological analysis (CFU/g of tissue) on Trypticase soy agar (TSA). Only those kidneys containing greater than 100 CFU/g of tissue were statistically analyzed. Pairwise comparisons were

conducted and differences were adjusted for multiple comparisons using the Tukey-Kramer method to maintain an overall alpha = .05 across all comparisons.

8.3 Results

8.3.1 Arginine Auxotrophy in Staphylococcus aureus

To examine arginine auxotrophy in *S. aureus*, eighty-two clinical *S. aureus* isolates collected from positive blood cultures at the University of Nebraska Medical Center were grown on Complete Defined Medium (CDM) with and without arginine. Similar to observations by Emmett and Kloos, only one *S. aureus* isolate (SA2126) had the ability to grow on CDM lacking arginine (CDM-R) following 48 h incubation, whereas all isolates grew on CDM containing arginine further confirming the arginine auxotrophic nature of *S. aureus* [8]. Furthermore, a community-associated *S. aureus* USA300 strain JE2 (Table 8.1) was unable to grow on CDMR following 48 h incubation at 37^{o}C. To extend these observations, JE2 was grown to stationary phase in CDM broth (5 x 10^{9} CFU) and plated on CDM-R and CDM lacking proline (CDM-P). Similar to the observations of Li and colleagues [18], *S. aureus* JE2 reverted to proline prototrophy at a rate of 1 x 10^{-6}; however, no colonies were isolated on CDM-R following five experimental attempts. Nevertheless, similar to observations by Gladstone, slight growth of JE2 was observed following five days of incubation in CDM-R broth [9]. These observations suggest that *S. aureus* has the inherent ability to synthesize arginine upon extended selection; however, the phenotype is not easily selected during growth in medium replete with amino acids.

8.3.2 Regulation of Arginine Biosynthesis by Carbon Catabolite Repression

Based on our observations that growth in CDM-R could be selected through extended incubation, we hypothesized that arginine biosynthesis was under transcriptional repression. Therefore, we screened a random *bursa aurealis* transposon library to isolate JE2 mutants able to grow on CDM-R. Two mutants were isolated that had the ability to grow on CDM-R; subsequent sequencing of the *bursa aurealis* insertions found they had inserted in *hprK* and *ccpA*. Both HprK and CcpA function to control carbon catabolite repression (CCR) in gram-positive bacteria [46]. Therefore, to completely eliminate CCR, a *ccpA* allelic replacement mutant was generated in JE2 through 80α transduction of the *ccpA::tetL* allele from MST14 (kind gift of M. Bischoff). As predicted, growth analysis in CDM-R demonstrated that JE2 *ccpA::tetL* enters exponential phase between 7–12 h and reaches a maximum OD_{600} of 4.5 after 24 h, whereas no growth was observed with wild type JE2 in CDM-R (Figure 8.1). Importantly, introduction of the *ccpA* complementation plasmid pNF266 abrogated growth of JE2 *ccpA::tetL*.

To further support the hypothesis that CCR functions to repress arginine biosynthesis, JE2 was grown in CDM-R lacking glucose but containing other, non-preferred carbon sources (Figure 8.2). Since CCR is alleviated when *S. aureus* is grown in media containing a non-preferred carbon source, it was hypothesized that JE2 would grow in CDM-R when glucose was replaced with a secondary carbon source. These experiments demonstrated that arabinose, sorbitol and pyruvate were able to support growth of JE2 when added to CDM-R (Figure 8.2). In contrast, glucose, fructose,

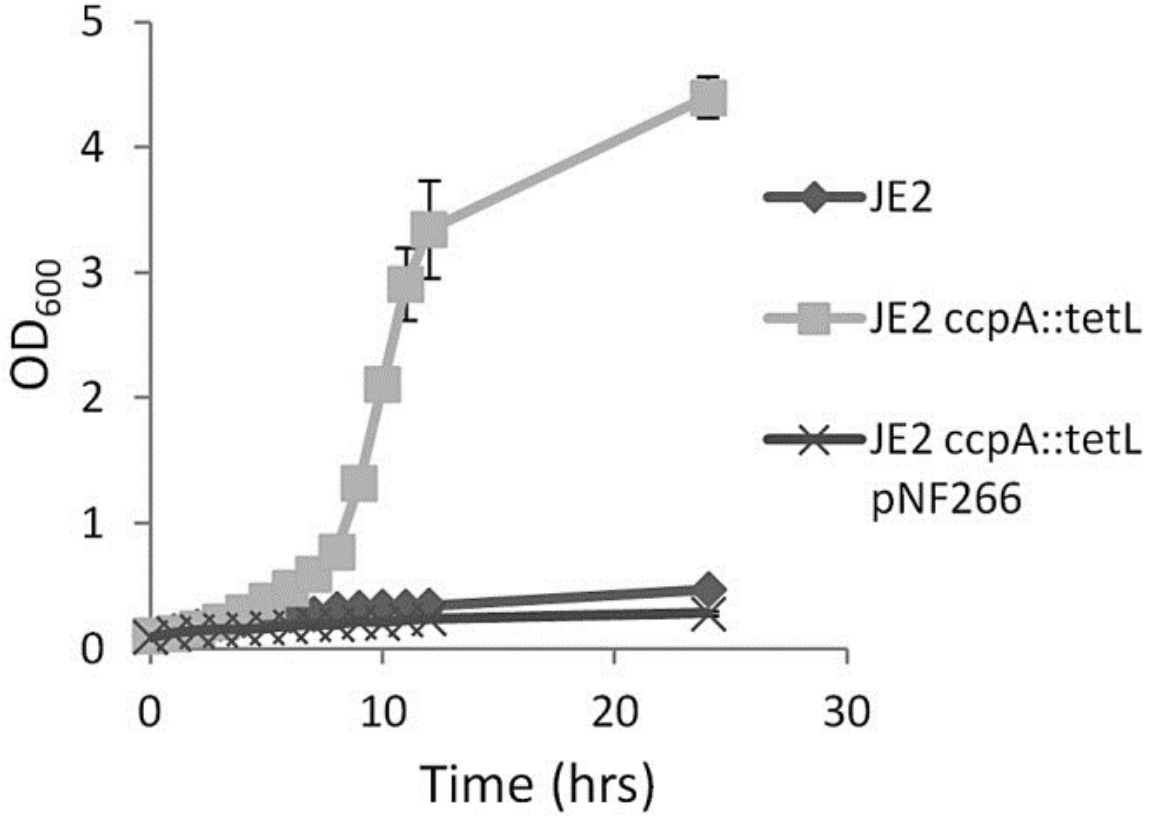


Figure 8.1. Interruption of *ccpA* facilitates growth in CDM-R. Growth analysis of JE2, JE2 *ccpA::tetL*, and JE2 *ccpA::tetL*/pNF266 (*ccpA* complement) in complete defined medium lacking arginine (CDM-R). Isolates were grown aerobically using a 10:1 flask to volume ratio. *S. aureus* strains containing a functional *ccpA* are unable to grow in CDM-R. Data represent means 6 SEM of three independent experiments.

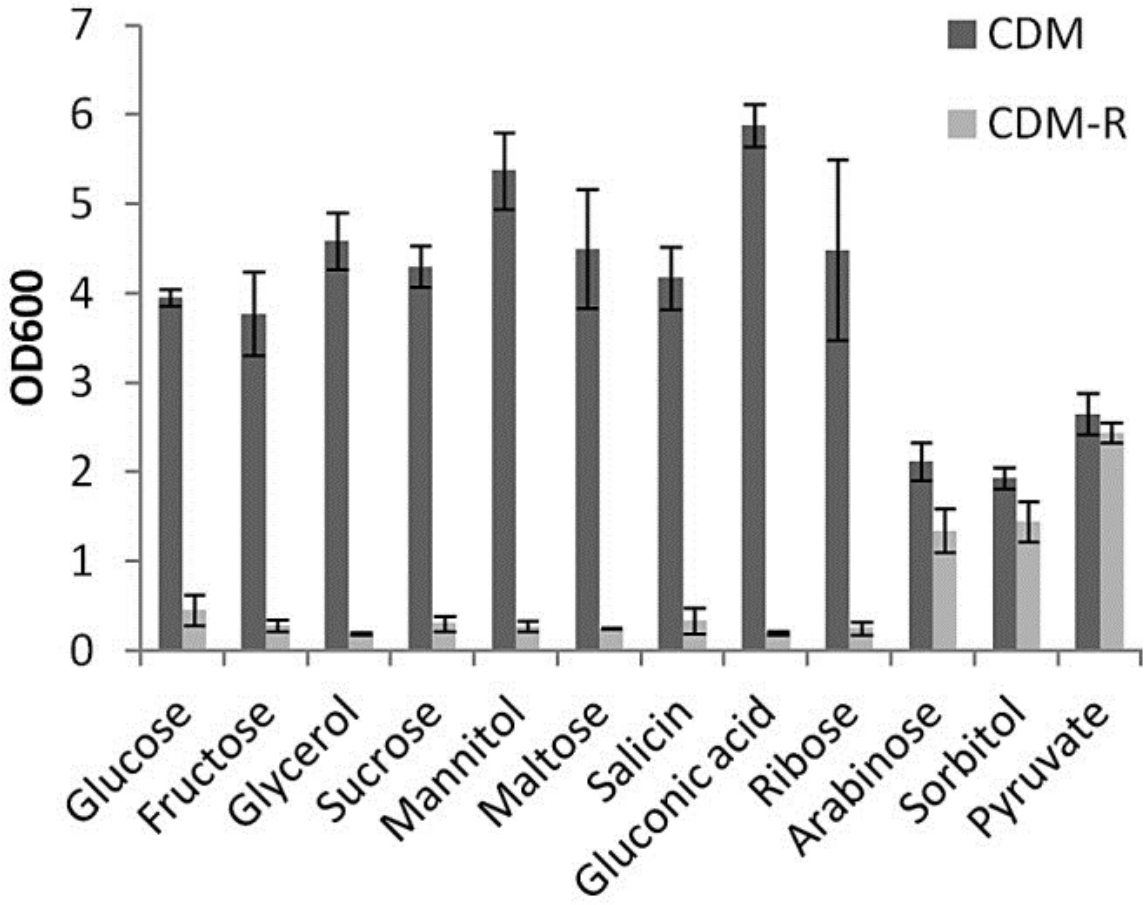


Figure 8.2. Growth of JE2 in CDM containing non-preferred carbon sources. JE2 was grown in CDM or CDM-R with the indicated carbon source at 37^{o}C for 18 hours. Significant growth in CDM-R was seen only when arabinose and sorbitol were used as carbon sources. Data represent means 6 SEM of five independent experiments.

glycerol, sucrose, mannitol, maltose, salicin, gluconic acid, and ribose were unable to support growth in CDM-R suggesting these carbohydrates do not derepress CcpA in JE2. In agreement with our results, Li and colleagues also determined that replacement of glucose with arabinose or sorbitol abrogated CcpA-mediated repression in *S. aureus* Newman and functioned to activate proline biosynthesis [18]. Overall, these data demonstrate that CCR functions to repress arginine biosynthesis, suggesting that arginine biosynthesis is linked to growth in niches where preferred carbon sources are limited.

8.3.3 Northern Analysis of argJBCDFGH in JE2 ccpA::ermB

Our preliminary data suggested that CCR functioned to repress an enzymatic step in the conserved arginine biosynthetic pathway via glutamate [47] (Figure 8.3). To further address this possibility, northern blot analysis was performed to address transcriptional expression of *argJBCDFGH* in JE2 *ccpA::ermB* in comparison to wildtype JE2. In *S. aureus*, *argDCJB* is arranged in an operon structure, whereas *argF* is transcribed as a monocistronic unit and *argGH* are co-transcribed. JE2 and JE2 *ccpA::ermB* were grown in CDM and CDM-R, respectively, to mid-exponential phase and mRNA was isolated. Utilizing DNA probes specific for each gene within the conserved pathway, *argDCJB* and *argF* expression was not detected in either JE2 or JE2 *ccpA::ermB* (Figure 8.4). However, although *argG* and *argH* transcripts were not detected in JE2, both transcripts were detected in JE2 *ccpA::ermB* (Figure 8.4). Therefore, although JE2 *ccpA::ermB* has the ability to grow on media lacking arginine, this strain does not appear to utilize the conserved arginine biosynthetic operon to

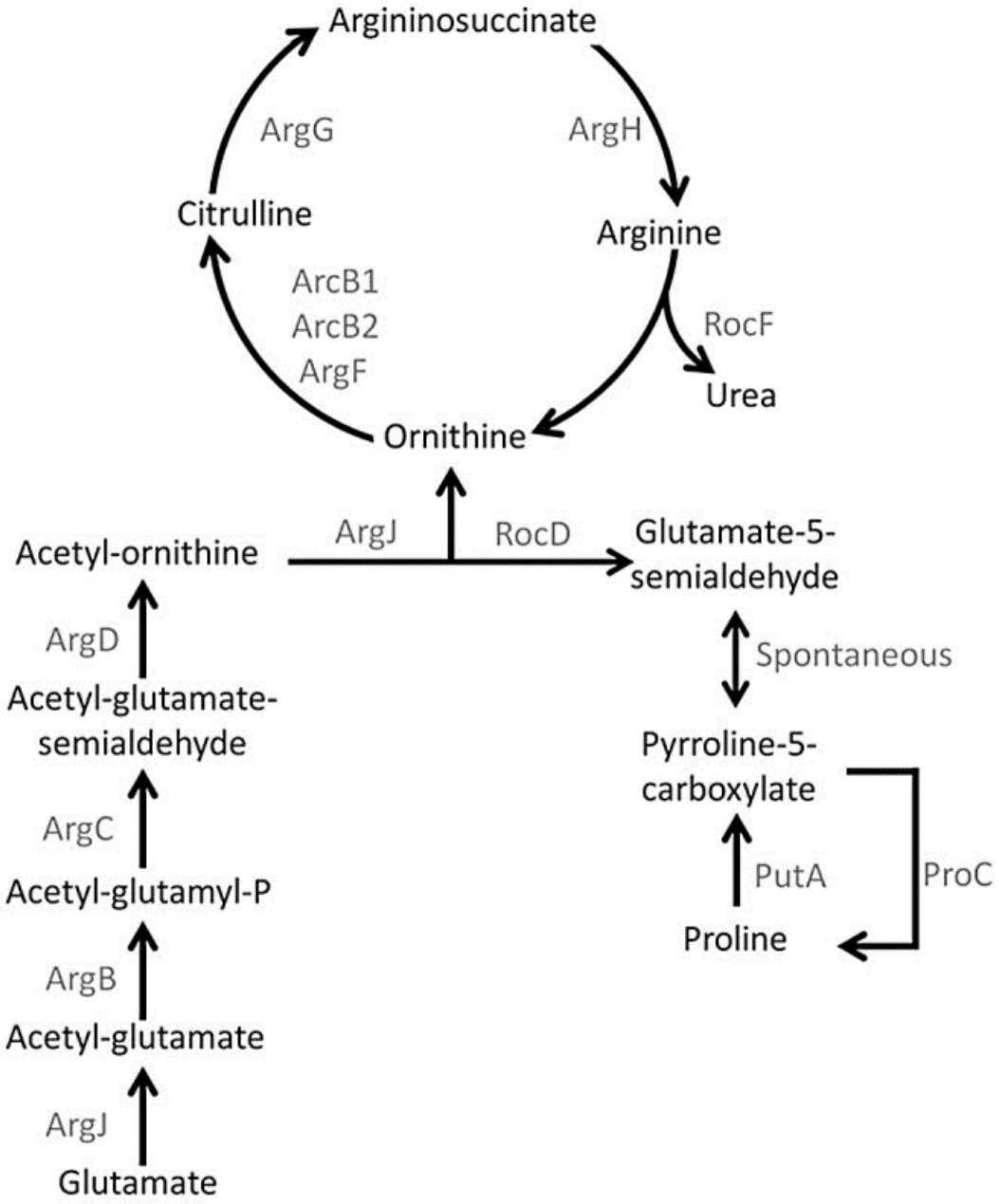


Figure 8.3. Arginine biosynthetic pathway via glutamate and proline. Figure depicts highly conserved arginine biosynthetic pathway via glutamate and the proposed pathway from proline via PutA, RocD, ArcB1, ArgG and ArgH. Note the previously established reverse pathway from arginine to proline via RocF, RocD and ProC.

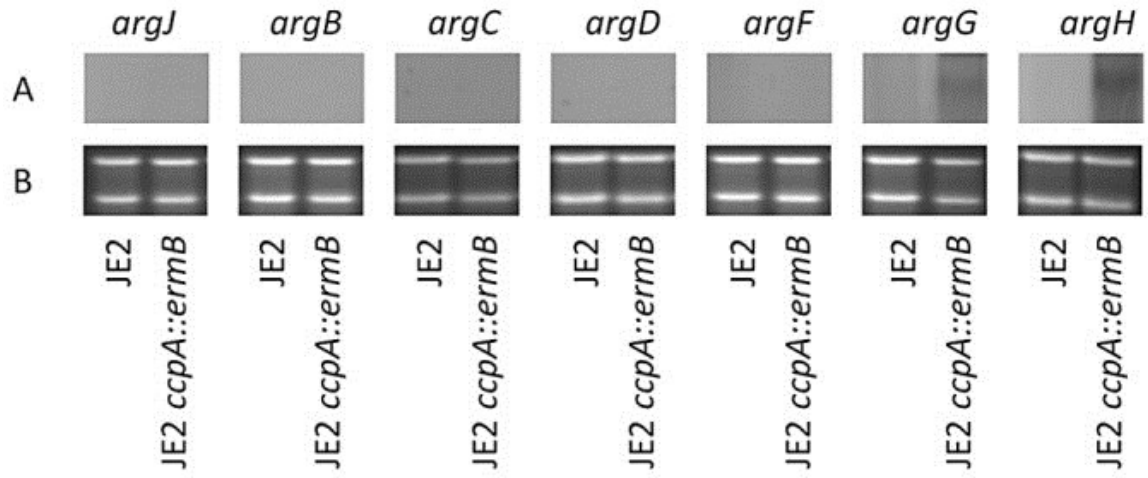


Figure 8.4. Northern analysis of arginine biosynthetic pathway in *S. aureus* JE2 and JE2 *ccpA::ermB*. JE2 and JE2 *ccpA::ermB* total RNA was isolated in mid-exponential phase of growth in CDM and CDM-R, respectively. DNA probes specific for *argJ*, *argB*, *argC*, *argD*, *argF*, *argG*, and *argH* were labeled with digoxygenin and detected using antidigoxigenin-AP Fab fragments (a). (b) shows 16s and 23s rRNA depicting equal RNA loading.

synthesize arginine in CDM-R. These results suggested the existence of a novel arginine biosynthetic pathway in *S. aureus*.

8.3.4 S. aureus Utilizes a Novel Proline Catabolic Pathway to Synthesize Arginine

Since our data indicated that glutamate was not the precursor for arginine synthesis in JE2 *ccpA::ermB*, other potential pathways were examined *in silico*. Based on the northern blot data demonstrating the expression of *argGH* in JE2 *ccpA::ermB*, we hypothesized that arginine may be synthesized via the urea cycle (Figure 8.3). *In silico* analysis predicted that either glutamate or proline have the potential to feed into the urea cycle to serve as substrates for arginine biosynthesis. To address this hypothesis, we examined amino acid consumption by JE2 and JE2 *ccpA::ermB* grown in CDM and CDM-R, respectively (Figure 8.5). These results demonstrated that both JE2 and JE2 *ccpA::ermB* consumed similar amounts of glutamate from the culture media following 24 h of growth. In contrast, JE2 *ccpA::ermB* consumed all available free proline from the culture medium, whereas only approximately 50% of the available free proline was consumed by JE2. Taken together, these observations allowed us to speculate that JE2 ccpA::ermB utilized proline via the urea cycle for arginine synthesis.

To further investigate this hypothesis, φ11 transducing lysates were prepared from defined JE2 *bursa aurealis* mutants with insertions in the following genes: *putA*, *proC*, *rocD*, *arcB1*, *arcB2*, *argF*, *argG*, *argH*, *argC*, *argB* and *argJ*. These *bursa aurealis* mutations (conferring erythromycin resistance) were transduced into JE2 *ccpA::tetL* and subsequently grown in CDM-R (Figure 8.6). Mutations in *argG*, *argH*, *putA*, *rocD*, and

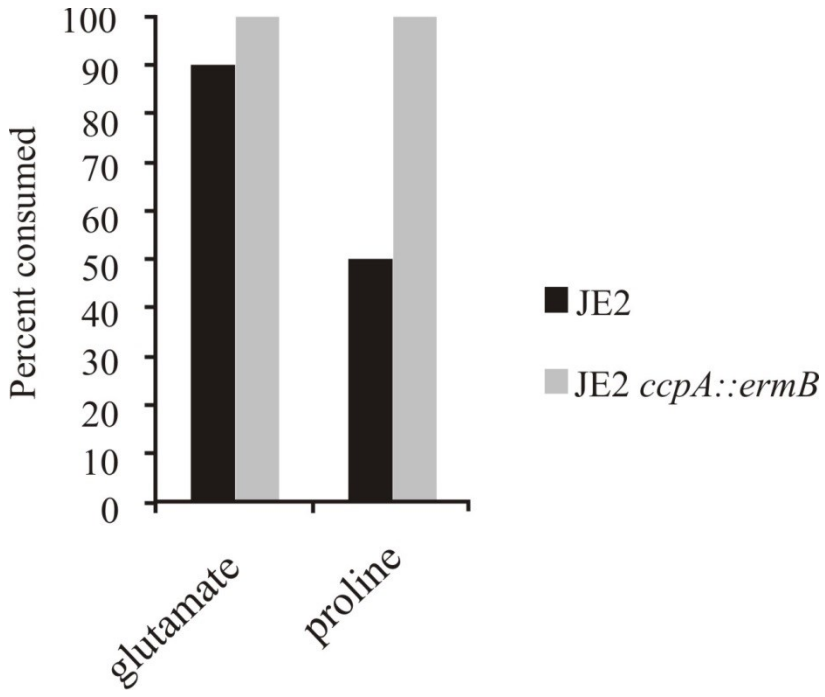


Figure 8.5. Amino acid analysis of JE2 and JE2 *ccpA::ermB* following growth in CDM. JE2 and JE2 *ccpA::ermB* were grown in CDM for 18 hours and supernatant was collected and analyzed for amino acid content. Percent of proline and arginine remaining is shown suggesting more efficient utilization of proline by JE2 *ccpA::ermB* in comparison to JE2.

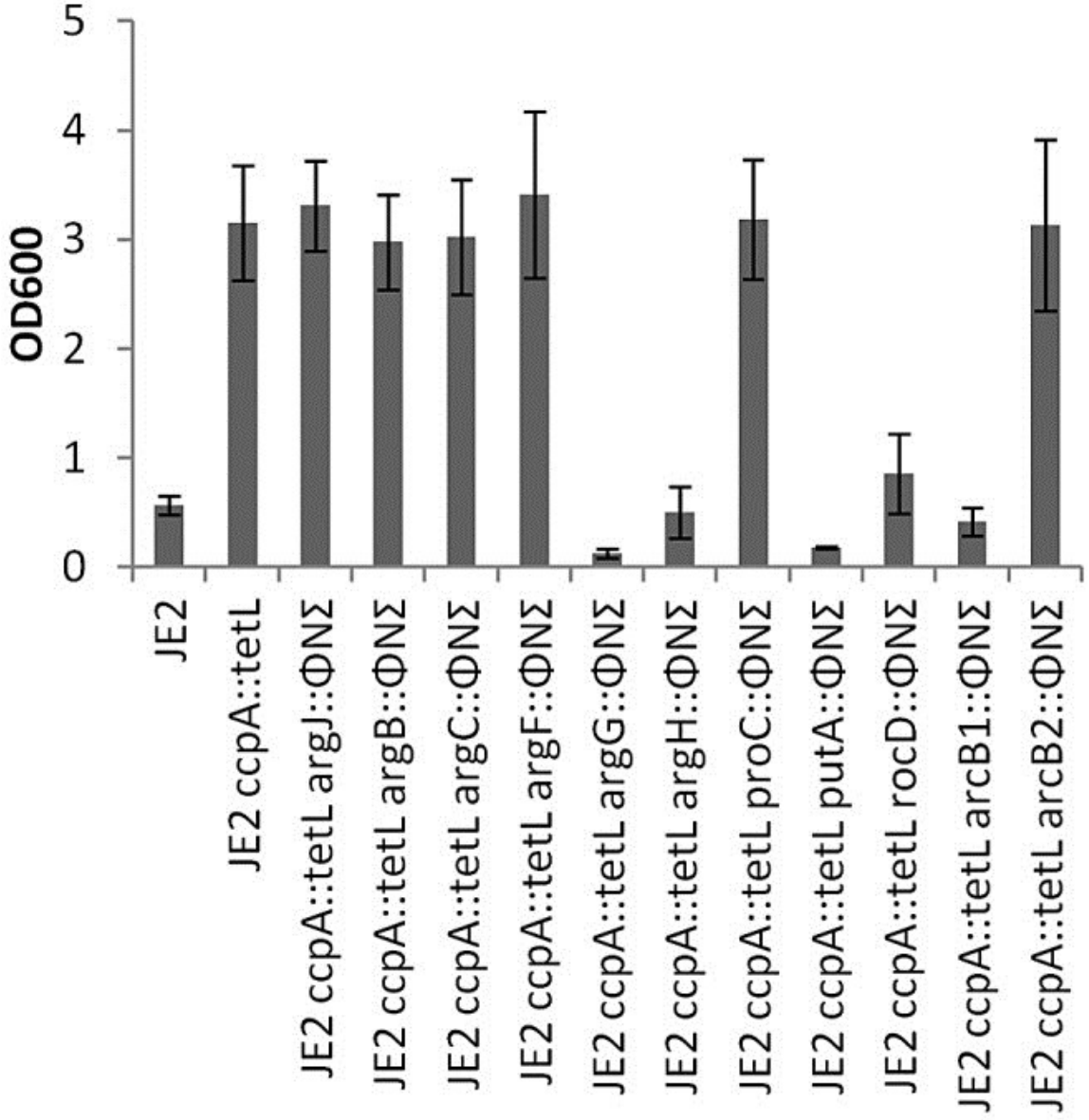


Figure 8.6. Determination of arginine biosynthesis-dependent genes in *S. aureus* JE2 *ccpA::tetL*. Defined *bursa aurealis* transposon mutants in *argJ*, *argB*, *argC*, *argF*, *argG*, *argH*, *proC*, *putA*, *rocD*, *arcB1*, and *arcB2* were transduced into JE2 *ccpA::tetL* and assessed for growth in CDM-R for 18 hours. Data represent means 6 SEM of three independent experiments.

arcB1 abrogated growth of JE2 *ccpA::tetL* in CDM-R. However, mutations in *argJ*, *argB*, *argC*, *argF*, *arcB2*, or *proC* had no effect on growth consistent with the prediction that arginine is synthesized from proline and not glutamate (Figure 8.6).

Two-dimensional (2D) ^{1}H-^{13}C heteronuclear single quantum coherence (HSQC) nuclear magnetic resonance (NMR) experiments were performed to confirm these data. JE2 and JE2 *ccpA::ermB* were grown in the presence of $^{13}C_5$-glutamate or $^{13}C_5$-proline in CDM and CDM-R, respectively. Based on our genetic studies, it was predicted that ^{13}C-labeled arginine would only be detected when JE2 *ccpA::ermB* was grown in CDM-R containing $^{13}C_5$-proline. As expected, ^{13}C-labeled arginine was detected when JE2 *ccpA::ermB* was propagated in the presence of $^{13}C_5$-proline but not with $^{13}C_5$-glutamate (Figure 8.7). Collectively, these results provide strong evidence that proline is the substrate for arginine biosynthesis in a *ccpA* genetic background. Furthermore, it is demonstrated that the highly conserved arginine biosynthetic pathway via glutamate is inactive under the growth conditions utilized in the study.

8.3.5 Arginine Auxotrophy in Other Staphylococcus aureus Strains

To determine whether our data regarding arginine biosynthesis were specific to the JE2 background, φ11 transducing lysates were prepared from JE2 *bursa aurealis argF* and *argH* mutants and introduced into RN4220 and Newman *ccpA* backgrounds. As previously noted with JE2 *ccpA::tetL*, an *argH* mutation abolished the ability of both RN4220 *ccpA::tetL* and Newman *ccpA::tetL* to grow in CDM-R, whereas a mutation in *argF* had no effect (Figure 8.8). Interestingly, RN4220 has the ability to grow in CDM-R

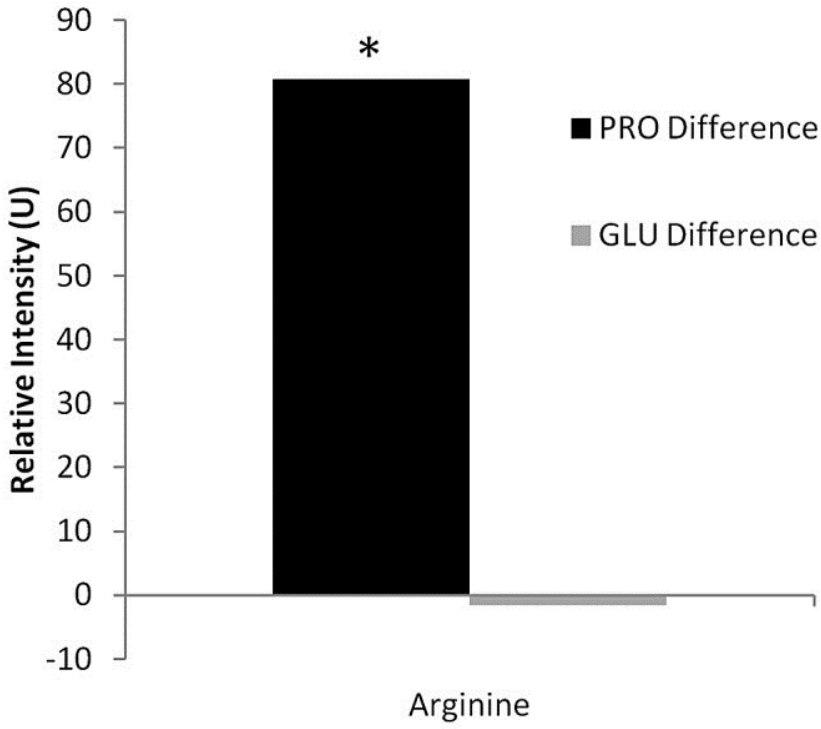


Figure 8.7. Two-dimensional (2D) ^{1}H-^{13}C heteronuclear single quantum coherence (HSQC) nuclear magnetic resonance (NMR) analysis of JE2 and JE2 *ccpA::ermB*. JE2 and JE2 *ccpA::ermB* were grown in the presence of $^{13}C_5$-glutamate or $^{13}C_5$-proline in CDM and CDM-R, respectively, and assayed using 2D ^{13}C HSQC NMR. The differences in ^{13}C-arginine relative intensity were determined by subtracting the average intensities between JE2 and JE2 *ccpA::ermB*, and a student's t-test was utilized to determine significance. A positive relative intensity value is indicative of a greater intensity of ^{13}C-arginine in JE2 *ccpA::ermB* in comparison to JE2. JE2 *ccpA::ermB* accumulated significantly greater amounts of ^{13}C-arginine when grown in CDM containing $^{13}C_5$-proline in comparison to JE2. Note that there was no significant difference in ^{13}C-arginine accumulation between JE2 and JE2 *ccpA::ermB* when grown in CDM and CDM-R, respectively, containing $^{13}C_5$-glutamate.

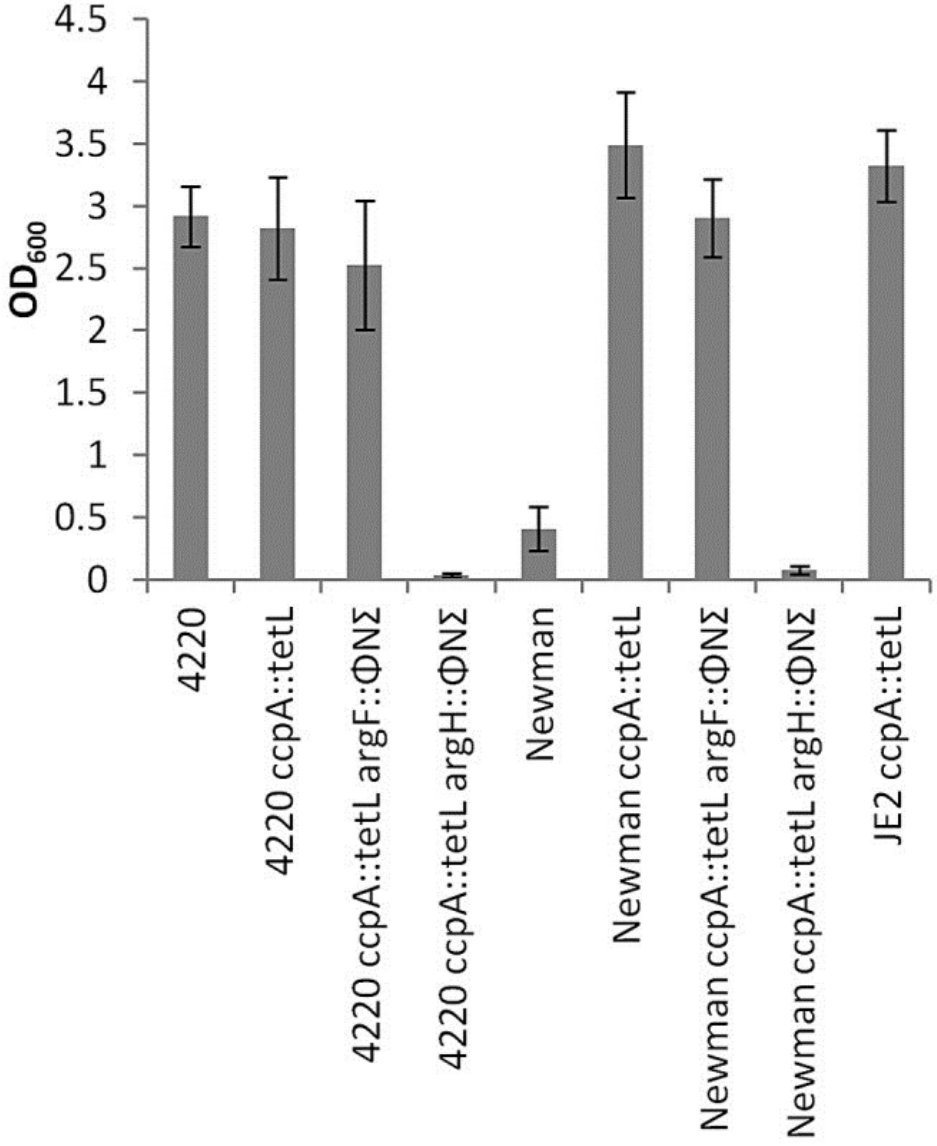


Figure 8.8. Determination of arginine biosynthesis-dependent genes in *S. aureus* Newman and RN4220. Defined *bursa aurealis* transposon mutants in *argF* and *argH* were transduced into Newman *ccpA::tetL* and RN4220 *ccpA::tetL* and assessed for growth in CDM-R for 18 hours. Data represent means 6 SEM of three independent experiments.

broth. Subsequent studies demonstrated that RN4220 reverted to arginine prototrophy at a frequency of 1.6 x 10^{-5}; however, sequence analysis of these mutants indicated they did not have mutations in *ccpA*, *hprK* or *ptsH* suggesting weak carbon catabolite repression in the RN4220 strain background. In addition, RN4220 *argH::*φNΣ was unable to grow in CDM-R broth whereas a *bursa aurealis* mutation in *argF* had no effect on growth suggesting RN4420 synthesizes arginine from proline but not from glutamate. Collectively, these data suggest that as a species, *S. aureus* has evolved to synthesize arginine via proline when growing in conditions lacking a preferred carbon source.

8.3.6 Virulence in a mouse kidney abscess model

C57BL/6 mice were inoculated retro-orbitally with 10^6 CFU of JE2, JE2 *argH::*φNΣ, JE2 *argC::*φNΣ, JE2 *putA::*φNΣ, or JE2 *argF::φNΣ*. The mice were harvested at 20 days and the kidneys were homogenized and CFU/gram of tissue determined (Figure 8.9). No statistical difference was determined between JE2 and JE2 *argF::*φNΣ, JE2 *argC*::φNΣ, or JE2 *putA*::φNΣ. However, a significant difference was noted between JE2 (mean $\log_{10}$ CFU of 5.31) and JE2 *argH*::φNΣ (mean $\log_{10}$ CFU of 4.21) indicating a potential function of *argH* and arginine biosynthesis in abscess development and persistence.

8.4 Discussion

The study of arginine biosynthesis has served as a paradigm for the regulon concept, originally coined by Maas and Clark, where the same transcriptional repressor

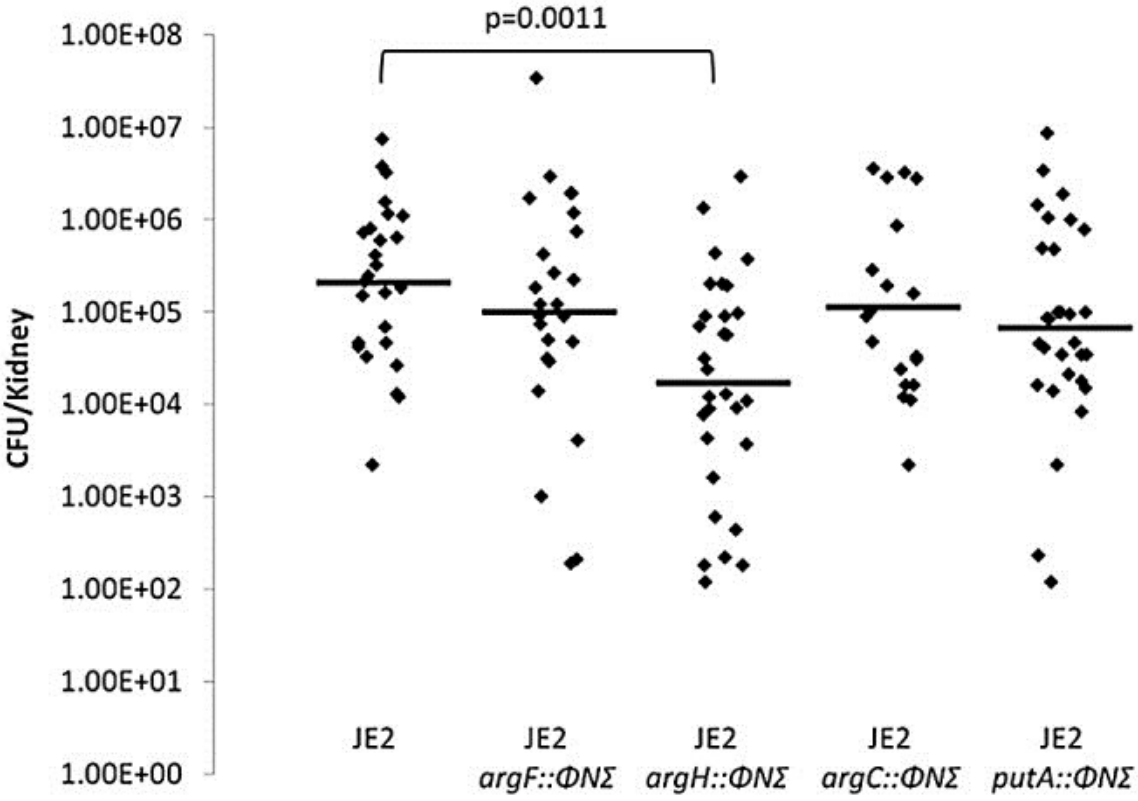


Figure 8.9. Mouse kidney abscess model. C57BL/6 mice were infected with 10^6 CFU of JE2 (n = 16 mice), JE2 *argF::*φNΣ (n = 14 mice), JE2 *argH::*φNΣ (n = 15 mice), JE2 *argC::*φNΣ (n = 13 mice), or JE2 *putA::*φNΣ (n = 18 mice). Kidneys were homogenized after 20 days and bacterial burden determined through viable count (CFU/gram tissue). Horizontal line represents median $\log_{10}$ CFU/gram; significant differences in bacterial burden were noted between JE2 and JE2 *argH::*φNΣ (**p<.01). Data were analyzed using two-way ANOVA.

regulates unlinked loci, ArgR [48]. Based on these studies and over 60 years of research, there are three established biochemical pathways, all utilizing glutamate as a substrate, that synthesize arginine in bacteria [47]. These three pathways primarily differ in the enzymes used to generate L-citrulline from N-acetyl-L-ornithine [47, 49]. All sequenced staphylococcal species analyzed to date encode the highly conserved ArgJBCDFGH on three separate unlinked transcriptional units; two operons (*argDCJB* and *argGH*) and one monocistronic gene (*argF*). Within the *S. aureus* USA300 FPR_3757 genome (NC_007793), the genes predicted to encode the arginine biosynthetic pathway are as follows: *argJ* (bifunctional ornithine acetyltransferase/glutamate N-acetyltransferase), SAUSA300_0185, EC 2.3.1.35/2.3.1.1; *argB* (acetylglutamate kinase), SAUSA300_0184, EC 2.7.2.8; *argC* (N-acetyl-gammaglutamyl-phosphate-reductase), SAUSA300_0186, EC 1.2.1.38; *argD* (acetylornithine transaminase), SAUSA300_0187, EC 2.6.1.11; *argF* (ornithine carbamoyl transferase), SAUSA300_1062, EC 2.1.3.3; *argG* (argininosuccinate synthase), SAUSA300_0864, EC 6.3.4.5; and *argH* (argininosuccinate lyase), SAUSA300_0863, EC 4.3.2.1. However, as previously reported and confirmed in this study, *S. aureus* is a functional arginine auxotroph when grown on complex laboratory media [8, 9]. In addition, no nonsense mutations or insertions were detected within the *argJBCDFGH* genes of the USA300 FPR_3757 genome or any other sequenced staphylococcal genome, suggesting that arginine biosynthesis is not a decaying pathway in the staphylococci. As discussed by Somerville and Proctor, in some cases, amino acid auxotrophies in *S. aureus* may be linked to TCA cycle inactivity or feedback inhibition due to growth in amino acid and glucose replete

media [50]. Our results are in agreement with this hypothesis where inactivation of *ccpA*, which represses the TCA cycle [51] and other genes that function to metabolize secondary carbon sources, was linked to arginine biosynthesis in *S. aureus*. In the presence of a preferred carbon source, the CcpA/Hpr complex represses a multitude of genes linked to central metabolism, amino acid metabolism and virulence [2, 52-54]. Therefore, based on previous studies, bioinformatic analyses of the *S. aureus* genome, and the work by Li and colleagues demonstrating that proline biosynthesis was linked to ccpA regulation, it was not unexpected to discover that arginine biosynthesis was linked to carbon catabolite repression [5, 18, 50, 55]. However, it was remarkable to discover that *S. aureus* does not use the conserved *argJBCDFGH* pathway to synthesize arginine via glutamate. Rather, we provide both genetic and biochemical evidence in support of a novel biosynthetic pathway, whereby *S. aureus* utilizes proline as a substrate via the urea cycle. First, mutations within *putA*, *rocD*, *arcB1*, *argG*, and *argH*, but not *argJ*, *argB*, *argC*, or *argF*, abolished growth of a *ccpA* mutant on CDM-R, providing genetic evidence that proline serves as a precursor for arginine synthesis (Figure 8.6). It is important to note that inactivation of *arcB1* abrogated growth of JE2 *ccpA::tetL* whereas a mutation in *arcB2* did not. *arcB1* (SAUSA300_2569) encodes the native ornithine carbamoyltransferase within the arginine deiminase operon whereas *arcB2* (SAUSA300_0062; ornithine carbamoyltransferase) is within the ACME pathogenicity island encoded arginine deiminase operon [56]. These data suggest that *ArcB1* and *ArcB2* are not functionally redundant or are not expressed under the same growth conditions. *arcB2* transcript is not detected using northern analysis (data not shown) under *in vitro*

growth conditions used in this study (CDM or CDM-R broth), however, it is unknown whether it is induced under other *in vivo* or *in vitro* growth conditions. Second, 2D ^{1}H-^{13}C HSQC NMR experiments provided compelling evidence that arginine is synthesized via proline and the urea cycle in a *S. aureus ccpA* mutant. Although there have been two reports demonstrating that proline is synthesized from arginine in *S. aureus* [17, 18], we are unaware of any reports demonstrating that arginine can be synthesized from proline. Li and colleagues demonstrated that CcpA binds to a *cre* site just upstream of *rocD*. Using the *cre* site from *pckA* as a consensus sequence [18], we identified potential *cre* sites upstream of *putA*, *arcB1*, and *argGH* (Figure 8.10). However, the function of these *cre* sites in regards to CcpA regulation has yet to be defined.

Previous studies have demonstrated that a *S. aureus ccpA* mutant also synthesizes proline from arginine via RocF (arginase), RocD (ornithine aminotransferase), and ProC (P5C reductase) [17, 18]. Collectively, these data and our observations suggest that under carbon-limiting conditions (*in vivo* environment), *S. aureus* can synthesize proline from arginine and arginine from proline depending on which amino acid is limited. Based on our findings and the existing literature, we propose a hypothetical model whereby free arginine is limited in the host during infection causing competition between the host and bacteria for arginine. In humans, L-arginine is a non-essential amino acid under homeostatic conditions. However, arginine becomes a ‘‘conditionally essential’’ [57] amino acid during sepsis or trauma due to its use as a substrate for inducible nitric oxide synthase [58] and function in cell-mediated immunity [59], protein synthesis [60] and wound healing [61, 62]. Indeed, recent studies have shown significant iNOS and arginase

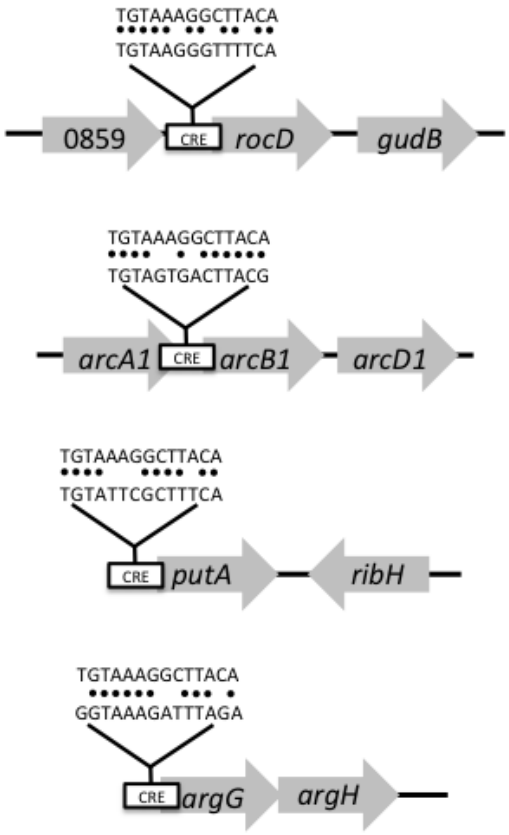


Figure 8.10. Putative *cre* sites in arginine biosynthesis-dependent genes. Using the *cre* site from *pckA* as a consensus sequence, putative *cre* sites were identified in *rocD*, *arcB1*, *putA*, *and argGH. cre* site from *pckA* is the top sequence whereas the putative *cre* site from the identified gene is the bottom sequence.

expression during *S. aureus* infection [63] (Figure 8.11), providing further support that a staphylococcal abscess may be an arginine-depleted environment based on the requirement of arginine for these host enzymes to function. In addition, low levels of L-arginine have been reported in plasma during sepsis [57], causing some investigators to suggest the use of L-arginine as a treatment modality [64]. Furthermore, arginine can serve as a substrate for arginine deiminase and subsequent direct ATP generation in the staphylococci [47].

Although little information is available regarding the concentration of free proline in a staphylococcal abscess, proline is the predominant amino acid found in collagen. Collagen is the most abundant protein in animals and type I collagen is a major constituent of the fibrotic wall surrounding staphylococcal abscesses [65] (Figure 8.11). Furthermore, *S. aureus* encodes two proteases, SspB and ScpA, which possess the ability to degrade collagen [66-68]. Therefore, our model predicts that *S. aureus* utilizes specific proteases to degrade collagen, resulting in the liberation of free proline or proline-containing peptides that are utilized to synthesize arginine via the urea cycle.

Strengthening this argument, earlier work demonstrated that mutants lacking the high affinity proline permease PutP are less virulent in animal models of infection [69, 70]. This proposed framework was initially tested using a mouse kidney abscess model previously utilized by Cheng and colleagues [71]. In this model, staphylococcal abscesses within the kidney are contained within a pseudocapsule-like structure; we hypothesized an *argH* and *putA* mutant would be attenuated in abscess persistence in comparison to wild type JE2, JE2 *argC*::φNΣ and JE2 *argF*::φNΣ due to the inability to utilize proline

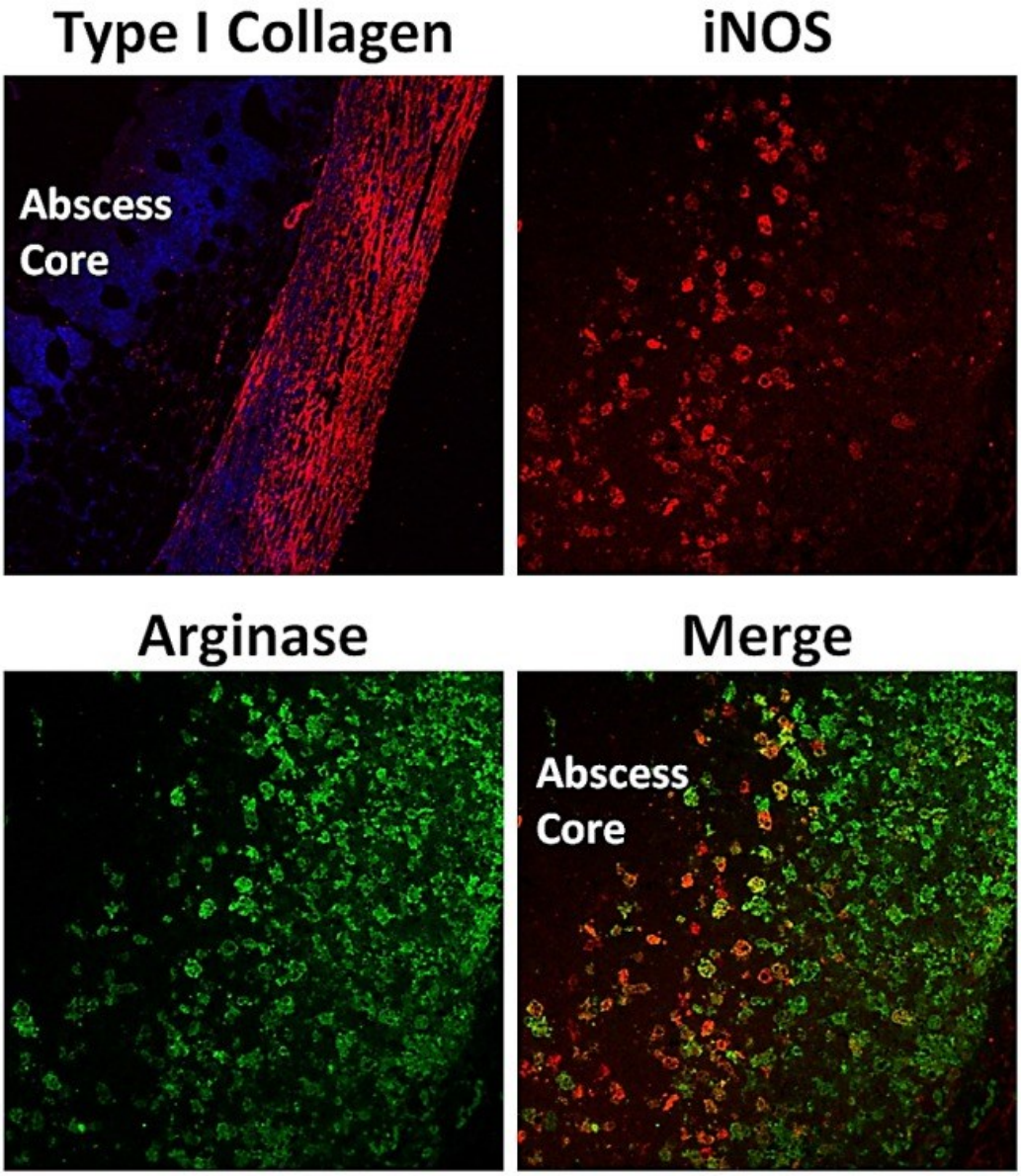


Figure 8.11. Mouse subcutaneous abscess model. C57BL/6 flank abscesses caused by subcutaneous injection of 5 x 10^5 CFU of *S. aureus* JE2. Tissues were processed for immunofluorescence staining for either type I collagen (red), inducible nitric oxide synthase (iNOS, red), or arginase (green). In addition to type I collagen, tissues were processed with the DAPI nuclear stain (blue) to accentuate the abscess core. Representative confocal microscopy images are presented for type I collagen (10 x magnification) and iNOS/arginase (20 x magnification).

from the pseudocapsule as a substrate for arginine synthesis. Supporting our model, in those kidneys containing staphylococcal abscesses, a significant 1 $\log_{10}$ difference was observed between JE2 and JE2 *argH*::φNΣ demonstrating the importance of arginine biosynthesis via the urea cycle in an in vivo infection model. As predicted, based on our *in vitro* data, no significant difference was observed between JE2, JE2 *argC*::φNΣ, and JE2 *argF*::φNΣ in the mouse kidney abscess model. However, in contrast to our predicted results, no significant difference in bacterial persistence was detected between JE2 and JE2 *putA*::φNΣ; PutA converts proline into pyrolline-5-carboxylate (Figure 8.3). It is known that the addition of either citrulline or ornithine to CDM-R can complement JE2 *ccpA::tetL putA*::φNΣ allowing growth. Therefore, it is possible that generation of citrulline or ornithine by arginine deiminase and ornithine carbamoyltransferase [47], respectively, circumvents and complements the proline requirement and facilitates the synthesis of arginine. In addition, since *argGH* is common to both the glutamate and proline pathways leading to the synthesis of arginine, an alternative interpretation of the data is that both pathways are active *in vivo* and have the ability to complement each other.

Finally, we have demonstrated that other *S. aureus* strains synthesize arginine from proline when CcpA activity is abolished, suggesting conservation of this pathway within the species. However, based on the conserved sequence analysis of the ArgJBCDFGH pathway within sequenced *S. aureus* isolates, we predict that this arginine biosynthetic pathway is active under growth conditions or niches that remain to be identified. Further work is required to dissect the evolving dogma regarding arginine

metabolism and the relationship between the host and *S. aureus* in the ‘‘war for arginine’’ during infection.

8.5 References

1. DeLeo, F.R., et al., *Community-Associated Meticillin-Resistant Staphylococcus aureus.* Lancet, 2010. **375**(9725): p. 1557-68.

2. Seidl, K., et al., *Staphylococcus aureus CcpA Affects Virulence Determinant Production and Antibiotic Resistance.* Antimicrobial Agents and Chemotherapy, 2006. **50**(4): p. 1183-94.

3. Majerczyk, C.D., et al., *Staphylococcus aureus CodY Negatively Regulates Virulence Gene Expression.* Journal of Bacteriology, 2008. **190**(7): p. 2257-65.

4. Pohl, K., et al., *CodY in Staphylococcus aureus: A Regulatory Link Between Metabolism and Virulence Gene Expression.* Journal of Bacteriology, 2009. **191**(9): p. 2953-63.

5. Somerville, G.A. and R.A. Proctor, *At the Crossroads of Bacterial Metabolism and Virulence Factor Synthesis in Staphylococci.* Microbiology and Molecular Biology Reviews : MMBR, 2009. **73**(2): p. 233-48.

6. Zhu, Y., et al., *RpiR Homologues May Link Staphylococcus aureus RNAIII Synthesis and Pentose Phosphate Pathway Regulation.* Journal of Bacteriology, 2011. **193**(22): p. 6187-96.

7. Zhu, Y., et al., *Tricarboxylic Acid Cycle-Dependent Attenuation of Staphylococcus aureus in vivo Virulence by Selective Inhibition of Amino Acid Transport.* Infection and Immunity, 2009. **77**(10): p. 4256-64.

8. Emmett, M. and W.E. Kloos, *Amino Acid Requirements of Staphylococci Isolated from Human Skin.* Canadian journal of Microbiology, 1975. **21**(5): p. 729-33.

9. Gladstone, G.P., *The Nutrition of Staphylococcus aureus; Nitrogen Requirements.* British Journal of Experimental Pathology, 1937. **18**(4): p. 322-333.

10. Kuroda, M., et al., *Whole Genome Sequencing of Meticillin-Resistant Staphylococcus aureus.* Lancet, 2001. **357**(9264): p. 1225-40.

11. Cunin, R., et al., *Biosynthesis and Metabolism of Arginine in Bacteria.* Microbiological Reviews, 1986. **50**(3): p. 314-52.

12. Vogel, R.H. and H.J. Vogel, *Acetylated Intermediates of Arginine Synthesis in Bacillus subtilis.* Biochimica et Biophysica Acta, 1963. **69**: p. 174-6.

13. Prozesky, O.W., *Arginine Synthesis in Proteus mirabilis.* Journal of General Microbiology, 1967. **49**(2): p. 325-34.

14. Rodriguez-Garcia, A., et al., *Characterization and Expression of the Arginine Biosynthesis Gene Cluster of Streptomyces clavuligerus.* Journal of Molecular Microbiology and Biotechnology, 2000. **2**(4): p. 543-50.

15. T., L., *Proline Biosynthesis*, in *Escherichia coli and Salmonella Typhimurium: Cellular and Molecular Biology*1987, ASM Press.: Washinton DC. p. 345-357.

16. Abdelal, A.T., *Arginine Catabolism by Microorganisms.* Annual Review of Microbiology, 1979. **33**: p. 139-68.

17. Townsend, D.E., et al., *Proline is Biosynthesized from Arginine in Staphylococcus aureus.* Microbiology, 1996. **142 (Pt 6)**: p. 1491-7.

18. Li, C., et al., *CcpA Mediates Proline Auxotrophy and is Required for Staphylococcus aureus pathogenesis.* Journal of Bacteriology, 2010. **192**(15): p. 3883-92.

19. Stulke, J. and W. Hillen, *Carbon Catabolite Repression in Bacteria.* Current Opinion in Microbiology, 1999. **2**(2): p. 195-201.

20. Deutscher, J., et al., *Protein Kinase-Dependent HPr/CcpA Interaction Links Glycolytic Activity to Carbon Catabolite Repression in Gram-Positive Bacteria.* Molecular Microbiology, 1995. **15**(6): p. 1049-53.

21. Henkin, T.M., *The Role of CcpA Transcriptional Regulator in Carbon Metabolism in Bacillus subtilis.* Fems Microbiology Letters, 1996. **135**(1): p. 9-15.

22. Hueck, C.J. and W. Hillen, *Catabolite Repression in Bacillus subtilis: A Global Regulatory Mechanism for the Gram-Positive Bacteria?* Molecular Microbiology, 1995. **15**(3): p. 395-401.

23. Schumacher, M.A., et al., *Structural Basis for Allosteric Control of the Transcription Regulator CcpA by the Phosphoprotein HPr-Ser46-P.* Cell, 2004. **118**(6): p. 731-41.

24. Deutscher, J. and M.H. Saier, Jr., *ATP-Dependent Protein Kinase-Catalyzed Phosphorylation of a Seryl Residue in HPr, A Phosphate Carrier Protein of the Phosphotransferase System in Streptococcus pyogenes.* Proceedings of the National Academy of Sciences of the United States of America, 1983. **80**(22): p. 6790-4.

25. Mijakovic, I., et al., *Pyrophosphate-Producing Protein Dephosphorylation by HPr kinase/phosphorylase: A Relic of Early Life?* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(21): p. 13442-7.

26. Bae, T., et al., *Staphylococcus aureus Virulence Genes Identified by Bursa aurealis Mutagenesis and Nematode Killing.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(33): p. 12312-7.

27. Kreiswirth, B.N., et al., *The Toxic Shock Syndrome Exotoxin Structural Gene is Not Detectably Transmitted by a Prophage.* Nature, 1983. **305**(5936): p. 709-12.

28. Duthie, E.S. and L.L. Lorenz, *Staphylococcal Coagulase; Mode of Action and Antigenicity.* Journal of General Microbiology, 1952. **6**(1-2): p. 95-107.

29. Kennedy, A.D., et al., *Epidemic Community-Associated Methicillin-Resistant Staphylococcus aureus: Recent Clonal Expansion and Diversification.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(4): p. 1327-32.

30. Pitz, A.M., et al., *Vancomycin Susceptibility Trends and Prevalence of Heterogeneous Vancomycin-Intermediate Staphylococcus aureus in Clinical Methicillin-Resistant S. aureus Isolates.* Journal of Clinical Microbiology, 2011. **49**(1): p. 269-74.

31. Hussain, M., J.G. Hastings, and P.J. White, *A Chemically Defined Medium for Slime Production by Coagulase-Negative Staphylococci.* Journal of Medical Microbiology, 1991. **34**(3): p. 143-7.

32. Climo, M.W., V.K. Sharma, and G.L. Archer, *Identification and Characterization of the Origin of Conjugative Transfer (oriT) and a Gene (nes) Encoding a Single-Stranded Endonuclease on the Staphylococcal Plasmid pGO1.* Journal of Bacteriology, 1996. **178**(16): p. 4975-83.

33. Horton, R.M., et al., *Gene Splicing by Overlap Extension: Tailor-Made Genes Using the Polymerase Chain Reaction.* BioTechniques, 1990. **8**(5): p. 528-35.

34. Bruckner, R., *Gene Replacement in Staphylococcus carnosus and Staphylococcus xylosus.* Fems Microbiology Letters, 1997. **151**(1): p. 1-8.

35. Sadykov, M.R., et al., *Tricarboxylic Acid Cycle-Dependent Regulation of Staphylococcus epidermidis Polysaccharide Intercellular Adhesin Synthesis.* Journal of Bacteriology, 2008. **190**(23): p. 7621-32.

36. Foster, T.J., *Molecular Genetic Analysis of Staphylococcal Virulence.* Methods in Microbiology, Vol 27, 1998. **27**: p. 433-454.

37. Charpentier, E., et al., *Novel Cassette-Based Shuttle Vector System for Gram-Positive Bacteria.* Applied and Environmental Microbiology, 2004. **70**(10): p. 6076-85.

38. Delaglio, F., et al., *NMRPipe: A Multidimensional Spectral Processing System Based on UNIX Pipes.* Journal of Biomolecular NMR, 1995. **6**(3): p. 277-93.

39. Johnson, B.A., *Using NMRView to Visualize and Analyze the NMR Spectra of Macromolecules.* Methods in Molecular Biology, 2004. **278**: p. 313-52.

40. Wishart, D.S., et al., *HMDB: A Knowledgebase for the Human Metabolome.* Nucleic Acids Research, 2009. **37**(Database issue): p. D603-10.

41. Cui, Q., et al., *Metabolite Identification via the Madison Metabolomics Consortium Database.* Nature Biotechnology, 2008. **26**(2): p. 162-4.

42. Akiyama, K., et al., *PRIMe: A Web Site that Assembles Tools for Metabolomics and Transcriptomics.* In Silico Biology, 2008. **8**(3-4): p. 339-45.

43. Okuda, S., et al., *KEGG Atlas Mapping for Global Analysis of Metabolic Pathways.* Nucleic Acids Research, 2008. **36**(Web Server issue): p. W423-6.

44. Caspi, R., et al., *The MetaCyc Database of Metabolic Pathways and Enzymes and the BioCyc Collection of Pathway/Genome Databases.* Nucleic Acids Research, 2010. **38**(Database issue): p. D473-9.

45. Zhang, B., et al., *NMR Analysis of a Stress Response Metabolic Signaling Network.* Journal of Proteome Research, 2011. **10**(8): p. 3743-54.

46. Sonenshein, A.L., *Control of Key Metabolic Intersections in Bacillus subtilis.* Nature Reviews. Microbiology, 2007. **5**(12): p. 917-27.

47. Lu, C.D., *Pathways and Regulation of Bacterial Arginine Metabolism and Perspectives for Obtaining Arginine Overproducing Strains.* Applied Microbiology and Biotechnology, 2006. **70**(3): p. 261-72.

48. Maas, W.K., *Studies on the Mechanism of Repression of Arginine Biosynthesis in Escherichia Coli. Ii. Dominance of Repressibility in Diploids.* Journal of Molecular Biology, 1964. **8**: p. 365-70.

49. Xu, Y., B. Labedan, and N. Glansdorff, *Surprising Arginine Biosynthesis: A Reappraisal of the Enzymology and Evolution of the Pathway in Microorganisms.* Microbiology and Molecular Biology Reviews : MMBR, 2007. **71**(1): p. 36-47.

50. Somerville, G.A., Proctor R.A., *Staphylococci in Human Disease*, in *The Biology of Staphylococci in Human Disease*, K.B. Crossley, Jefferson, K. K., Archer, G. L., Fowler, V.G., Editor 2009, Blackwell Publishing Ltd.: Oxford. p. 3-18.

51. Sadykov, M.R., et al., *CcpA Coordinates Central Metabolism and Biofilm Formation in Staphylococcus epidermidis.* Microbiology, 2011. **157**(Pt 12): p. 3458-68.

52. Seidl, K., M. Bischoff, and B. Berger-Bachi, *CcpA Mediates the Catabolite Repression of tst in Staphylococcus aureus.* Infection and Immunity, 2008. **76**(11): p. 5093-9.

53. Seidl, K., et al., *Staphylococcus aureus CcpA Affects Biofilm Formation.* Infection and Immunity, 2008. **76**(5): p. 2044-50.

54. Seidl, K., et al., *Effect of a Glucose Impulse on the CcpA Regulon in Staphylococcus aureus.* BMC Microbiology, 2009. **9**: p. 95.

55. Becker, S.A. and B.O. Palsson, *Genome-Scale Reconstruction of the Metabolic Network in Staphylococcus aureus N315: An Initial Draft to the Two-Dimensional Annotation.* BMC Microbiology, 2005. **5**: p. 8.

56. Diep, B.A., et al., *The Arginine Catabolic Mobile Element and Staphylococcal Chromosomal Cassette mec Linkage: Convergence of Virulence and Resistance in the USA300 Clone of Methicillin-Resistant Staphylococcus aureus.* The Journal of Infectious Diseases, 2008. **197**(11): p. 1523-30.

57. Davis, J.S. and N.M. Anstey, *Is Plasma Arginine Concentration Decreased in Patients with Sepsis? A Systematic Review and Meta-Analysis.* Critical Care Medicine, 2011. **39**(2): p. 380-385.

58. Fang, F.C., *Perspectives Series: Host/Pathogen Interactions. Mechanisms of Nitric Oxide-Related Antimicrobial Activity.* The Journal of Clinical Investigation, 1997. **99**(12): p. 2818-25.

59. Ochoa, J.B., et al., *Effects of L-Arginine on the Proliferation of T lymphocyte Subpopulations.* JPEN. Journal of Parenteral and Enteral Nutrition, 2001. **25**(1): p. 23-9.

60. Luiking, Y.C., et al., *The Role of Arginine in Infection and Sepsis.* JPEN. Journal of Parenteral and Enteral nutrition, 2005. **29**(1 Suppl): p. S70-4.

61. Shi, H.P., et al., *Supplemental L-Arginine Enhances Wound Healing in Diabetic Rats.* Wound Repair and Regeneration : Official Publication of the Wound Healing Society [and] the European Tissue Repair Society, 2003. **11**(3): p. 198-203.

62. Witte, M.B. and A. Barbul, *Arginine Physiology and its Implication for Wound Healing.* Wound Repair and Regeneration : Official Publication of the Wound Healing Society [and] the European Tissue Repair Society, 2003. **11**(6): p. 419-23.

63. Thurlow, L.R., et al., *Staphylococcus aureus Biofilms Prevent Macrophage Phagocytosis and Attenuate Inflammation in vivo.* Journal of Immunology, 2011. **186**(11): p. 6585-96.

64. Kalil, A.C., *Is it Time to Replace L-Arginine in Severe Sepsis?* Critical Care Medicine, 2011. **39**(2): p. 417-418.

65. Shoulders, M.D. and R.T. Raines, *Collagen Structure and Stability.* Annual Review of Biochemistry, 2009. **78**: p. 929-958.

66. Kantyka, T., L.N. Shaw, and J. Potempa, *Papain-Like Proteases of Staphylococcus Aureus.* Cysteine Proteases of Pathogenic Organisms, 2011. **712**: p. 1-14.

67. Ohbayashi, T., et al., *Degradation of Fibrinogen and Collagen by Staphopains, Cysteine Proteases Released from Staphylococcus aureus.* Microbiology-Sgm, 2011. **157**: p. 786-792.

68. Rice, K., et al., *Description of Staphylococcus Serine Protease (ssp) Operon in Staphylococcus aureus and Nonpolar Inactivation of sspA-Encoded serine Protease.* Infection and Immunity, 2001. **69**(1): p. 159-169.

69. Bayer, A.S., et al., *Impact of the High-Affinity Proline Permease Gene (putP) on the Virulence of Staphylococcus aureus in Experimental Endocarditis.* Infection and Immunity, 1999. **67**(2): p. 740-744.

70. Coulter, S.N., et al., *Staphylococcus aureus Genetic Loci Impacting Growth and Survival in Multiple Infection Environments.* Molecular Microbiology, 1998. **30**(2): p. 393-404.

71. Cheng, A.G., et al., *Genetic Requirements for Staphylococcus aureus Abscess Formation and Persistence in Host Tissues.* Faseb Journal, 2009. **23**(10): p. 3393-3404.

CHAPTER 9

DISPENSABILITY OF D-ALANINE RACEMASE IN *MYCOBACTERIUM SMEGMATIS* AND *MYCOBACTERIUM TUBERCULOSIS* UNDERLIES A NOVEL PATHWAY OF D-ALANINE BIOSYNTHESIS

9.1 Introduction

Tuberculosis (TB) is a major global health problem and still the second worldwide cause of death from an infectious disease. Critical challenges to achieve the global targets for TB control and elimination include drug resistant TB, the co-epidemics of TB and HIV and the lack of more effective tools to prevent, diagnose and treat the disease [1]. The cell wall of *Mycobacterium tuberculosis*, the main etiological agent of human TB, plays an important role in mycobacterial pathogenesis and is a major source of targets for design of antimicrobial agents and attenuated mutants [2-4]. As in other bacteria, peptidoglycan strands and their crosslinks are critical components and stabilizers of the mycobacterial cell wall backbone. D-alanine is an essential building block of mycobacterial peptidoglycan and other eubacteria as it is involved in peptidoglycan crosslink formation [4-9].

The main source of D-alanine is the conversion of L-alanine by D-alanine racemase (Alr), an enzyme encoded by the *alr* gene [4, 7, 9]. The critical role of D-alanine in the cell wall peptidoglycan highlights the importance of its biosynthetic pathway(s) as targets for generation of new antimicrobial agents and attenuated mutants. Core to that design is the determination of *alr* essentiality and its role in the context of mycobacterial pathogenesis. Studies on *alr* in the fast growing species *Mycobacterium*

smegmatis, rendered apparently contradictory results. In our initial report, we found that the *M. smegmatis alr* gene was not required in the absence of D-alanine in the recommended Middlebrook 7H9 mycobacterial medium [10]. Our data also supported the existence of an alternative pathway of D-alanine biosynthesis in this species. Contradictory results were reported as *M. smegmatis* growth was assessed in Luria-Bertani or Middlebrook 7H11 media [11] since *alr* mutants were found to be auxotrophic for D-alanine. Moreover, recent reports in *M. tuberculosis* H37Rv suggest that *alr* mutants are also auxotrophic for D-alanine in both Middlebrook 7H9 and 7H11 [12]. Thus, according to these two reports, the *alr* gene is required in the absence of D-alanine in both *M. smegmatis* and *M. tuberculosis*. In that context, these latter results negate the existence of an alternative endogenous pathway of D-alanine biosynthesis in both species, as D-alanine would be an absolute requirement for growth of *alr* mutant strains.

Resolution of the controversy surrounding *alr* essentiality in mycobacteria is important, not only in the context of the role of Alr in mycobacteria, but also for its implications in the design of more effective drugs and vaccines targeting the endogenous synthesis of D-alanine in *M. tuberculosis* strains found in clinical settings. Moreover, evidence of horizontal gene transfer in mycobacteria [13, 14], highlights the impact that the potential transfer of genes involved in an alternative pathway of D-alanine biosynthesis might have in TB epidemiology and management. Herein, we describe the results of our studies aiming at determining *alr* essentiality in *M. smegmatis* and *M. tuberculosis* and possible implications for future *in vivo* studies.

9.2 Materials and Methods

9.2.1 Bacterial strains, plasmids, phages, culture conditions and nucleic acid manipulation.

Bacterial strains, plasmids and phages used in this study are listed in Table 9.1. *E. coli* and, in some cases *M. smegmatis* strains, were grown at 37ºC in broth or agar (1.5%, Difco Laboratories, Detroit, Mich.) using LB broth base, containing Tryptone, a pancreatic digest of casein (BD, Franklin Lakes, NJ). For *M. smegmatis* strains, LBT medium was LB broth or agar supplemented with 0.05% Tween 80 (Fisher Scientific, Pittsburgh, PA). For routine growth, *M. smegmatis* strains were cultured at 37ºC in Middlebrook 7H9 base broth (BD) or agar (Fisher Scientific) supplemented with ADC (0.5% bovine serum albumin fraction V (Fisher Scientific), 0.01 M glucose, 0.015 M sodium chloride, 0.2% glycerol, and for broth media only 0.05% Tween-80 ; MADC). *M. tuberculosis* strains were grown in MADC broth or 7H9 based agar media supplemented with ADC or OADC (ADC plus 0.6 mg/mL oleic acid). In some experiments, the following supplements (Sigma Chemical Co., St. Louis, MO) were individually added at specified concentrations: L-serine, D-serine, L-glutamic acid, D-glutamic acid, L-arginine, L-tryptophan, L-asparagine, L-aspartic acid and L-lysine, D-alanine, and pyruvate. Antimicrobial agents used were ampicillin (0.05 mg/mL); kanamycin A monosulfate (0.025 mg/mL), hygromycin B (0.075 to 0.1 mg/mL; Roche Molecular

Table 9.1. Bacterial strains, plasmids and phages used in this study.

Bacterial strains and vectors	Description	Reference
E. coli DH5α	F$^-$ *lacZDM15 endA1 hsdR17 supE44 gyrA96 relA1*	Invitrogen Life Technologies
M. smegmatis mc^{2}155	Alr$^+$; *M. smegmatis* ATCC 607 mutant, high transformation efficiency	[15]
M. smegmatis TAM23	Alr$^-$; Kan^r; *M. smegmatis alr* insertion mutant derived from *M. smegmatis* mc^{2}155	[10]
M. smegmatis TAM23-12	Alr$^-$; Kan^r; Hig^r; *M. smegmatis alr* and *murI* double mutant derived from *M. smegmatis* TAM23, D-alanine auxotroph	This study
M. smegmatis GPM 292- GPM295	Alr-$^-$; Kan^r; *M. smegmatis alr* deletion mutant derived from *M. smegmatis* mc^{2}155	This study
E. coli DH5α pBUN307	Kan^r. *E. coli* DH5α transformed with pBUN307	This study
M. tuberculosis H37Rv	Laboratory reference strain.	ATCC 25,618
M. tuberculosis H37Rv (pBUN307)	Kan^r. *alr* merodiploid strain. *M. tuberculosis* H37Rv transformed with pBUN307	This study
TBcol1	Kan^r, Hyg^r *alr* chromosomal mutant of *M. tuberculosis* H37Rv transformed with pBUN307	
pBUN73	Kan^r Tet^r. pMV203 derivative	Barletta lab. Unpublished
pBUN82	Kan^r. *M. smegmatis* wild type *alr* gene cloned into pMV262	[16]
pMV262	Replicating *E. coli-Mycobacterium* shuttle plasmid; carries P*hsp60* promoter upstream from polylinker site	[17]
pBUN201	pCR2.1 derivative carrying *M. tuberculosis alr* gene	This study
pBUN205	pET15b derivative carrying *M. tuberculosis alr* gene	This study
pCR2.1	Original TA cloning vector	Invitrogen Life Technologies
pET15b	Standard Expression vector	Novagen
pBUN307	Kan^r. Replicative shuttle plasmid carrying *M. tuberculosis* H37Rv *alr* transcribed under the control of its own promoter	This study
pBUN369	Tet^r Replicative shuttle plasmid. pBUN73 derivative.	This study
pBUN373	Tet^r Replicative shuttle plasmid, pBUN369 derivative carrying the *M. smegmatis alr* gene from pBUN82	This study
pBUN381	pBlueScript KS+ derivative carrying. *M. tuberculosis alr gene* flanking regions in a 3.8 kb fragment	This study
pTAMU3	Amp^r Hyg^r; pYUB412 derivative carrying *M. smegmatis alr* gene	[10]
pYUB412	Amp^r Hyg^r; *E. coli-Mycobacterium* integration-proficient vector	[18]

Table 9.1. (Continued)

Bacterial strains and vectors	Description	Reference
phAE159(Tn*5371*)	Derivative of phAE159 carrying transposon Tn*5371* with a Hyg^r marker	[19]
phAE87	Thermosensitive phasmid. PH101 derivative carrying cosmid pYUB328	[20]
pYUB854	Hyg^r Cosmid. Designed for phasmid construction and directional cloning of homologous recombination substrates	[21]
pBUN287	Hyg^r. pYUB854 derivative. Carries the 5'end of *alr* gene from *M. tuberculosis* H37Rv	This study
pBUN288	Hyg^r. pBUN287 derivative. Carries the 5'and 3'ends of *alr* gene from *M. tuberculosis* H37Rv and its adjacent regions	This study
phBUN1	Hyg^r phasmid. phAE87 derivative with pYUB328 replaced by pBUN288	This study

Biochemicals, Indianapolis, Ind.); D-cycloserine (0.00117 to 0.150 mg/mL); and tetracycline (0.0125 mg/mL).

M. smegmatis growth curves were determined in minimal broth medium as previously described [10]. *M. smegmatis* and *E. coli* broth cultures were incubated with shaking at 200 rpm in an Innova 4300 rotary incubator (New Brunswick Scientific Co. Inc., Edison, N.J.). *M. tuberculosis* broth cultures were incubated horizontally in standing flasks. Phage stocks were prepared following standard procedures and stored in MP buffer made with 50 mM Tris-HCl pH 7.5 (MP Biomedicals, LLC. Aurora, OH, USA), 150 mM NaCl (Merck KGaA, Danmstadt, Germany), 2 mM $CaCl_2$ (Sigma-Aldrich, CO. St. Louis, MO, USA), and 10 mM $MgSO_4$-$7H_2O$ (MP Biomedicals, LLC. Aurora, OH, USA), following standard procedures [22]. Specific protocols are further described.

Standard procedures were followed for nucleic acid purification, enzyme restriction digestions, ligations, and agarose gel electrophoreses [23]. Bacterial transformations and transfections were performed as described [22].

9.2.2 Construction, verification and complementation of M. smegmatis alr deletion mutant.

The *alr* flanking regions were amplified from 50 ng of *M. smegmatis* mc^2155 DNA using primer sets AlrK01/AlrK02 and AlrK03/AlrK04, as described [11] (Table 9.2). The resulting fragments (1.2 kb and 1.3 kb respectively) were subcloned into pBluescript KS+ to generate pBUN381. A 3.8 kb fragment from pBUN381, carrying the

pUC18K *aphA-3* kanamycin resistance marker outflanked by the *alr* uspstream and downstream regions, was used to transform *M. smegmatis* mc^2155. Transformants GPM 292, GPM293, GPM294 and GPM295 were selected on LBT supplemented with 5.0 mM D,L-alanine. Deletion of *alr* was verified by PCR and Southern blotting analysis.

For PCR verification, *M. smegmatis* genomic DNA was amplified with the following primer sets: AlrInt1F/AlrInt1R and AlrIntF /AlrIntR, targeting the deleted *alr* region, as well as with primers 5alr2 and 3alr2, targeting the wild *alr* gene flanking sequences. For Southern blotting analysis, approximately 3.0 µg of *Sma*I digested *M. smegmatis* wild type and mutant genomic DNA were transferred to nylon membranes and hybridized with a radiolabelled DNA fragment amplified from wild type mc^2155 DNA using primers AlrK01 and AlrK02. Prehybridization, hybridization and high-stringency washes were done at 65ºC.

To complement the *alr* deletion mutant strain GPM292, cells were grown to mid-exponential phase in LBT supplemented with 5.0 mM D,L-alanine. Cells were transformed by electroporation with pTAMU3 as previously described [10]. Transformants were selected on LBT agar supplemented with 5.0 mM D,L-alanine and 0.1 mg/ml hygromycin B (A.G. Scientific, San Diego, CA) until normal sized colonies were visually observed at 37ºC. Individual transformant colonies were grown in 1.0 mL of LBT supplemented with 5.0 mM D,L-alanine and 0.1 mg/mL hygromycin B, and cells from 0.001 mL growing cultures were directly lysed in the thermocycler and subjected to PCR amplification to verify the presence of the hygromycin resistance gene from pTAMU3. Amplification was carried out under standard conditions with primers Hygro-

2F and Hygro-2R, using the following settings: 95°C for 5 minutes, followed by 30 cycles of 95°C for 30 sec, 55°C for 30 seconds, 72°C for 30 sec, and a final extension at 72°C for 7 minutes.

9.2.3 D-Alanine racemase enzymatic activity assays.

M. smegmatis mc^2155, TAM23 and GPM292 cells were grown to mid-exponential phase (OD_{600nm} ca. 0.6-0.7) in 200 mL LBT supplemented with 1.0 mM D-alanine, harvested, and concentrated 20-fold in 50 mM Tris-HCl (pH 8.0). Cell extracts were prepared as described [16], except that a Centriprep YM10 (Millipore, Billerica, MA) concentration step was added prior to final filtration and dialysis. Protein concentrations were determined by a modified Lowry protein assay (http://www.ruf.rice.edu/~bioslabs/methods/protein/lowry.html).

Alr reactions were performed in triplicate in a final volume of 50 µL containing 50 mM Tris-HCl (pH 8.0), 0.1 mM pyridoxal phosphate, 15 mM D-alanine, 0.5 µCi ^{14}C D-alanine, and approximately 100 µg of each cell extract. Reactions were incubated at 37ºC and in a water bath for 15 min, and stopped by adding SDS to a final concentration of 0.25%. Parallel negative controls were performed with heat inactivated extracts otherwise processed in identical manner. Standards for ^{14}C-D-alanine, ^{14}C-L-alanine, and a 50% D, L-alanine mixture were prepared with the same reaction components, heat inactivated mc^2155 extract, and 0.5 µCi of total ^{14}C-alanine. Chiral plates (Macherey-Nagel, Bethlehem, PA; 10 x 20 cm) were baked for 30 minutes at 100 ºC prior to sample loading. Reaction mixtures (5 µL) were spotted on the plate, 2.5 µL at a time, allowing

the spots to dry between applications. Plates were developed by ascending chromatography in a chamber saturated with vapors of the running solvent (acetone:methanol:water 10:2:2) until the solvent front reached 0.5 cm from the top of the plate.

Plates were dried, wrapped in plastic, and exposed to a Kodak Imaging Screen-K for 4 h, and analyzed using a Biorad Molecular Imager PharosFX and Quantity One Software. To quantify spots, the rectangular volume analysis tool was used. For overlapping oval spots (e.g., Alr activity in mc^2155), the rectangular box for a particular spot was extended up to the inflection point in the spot images. Global subtraction was used to reduce background.

9.2.4 Growth of M. smegmatis alr mutants in liquid media.

Minimal concentrations of D-alanine required for the mutant strains to yield significant growth were determined in both LBT and MADC broth, supplemented with 0.1 mM – 1.0 mM D-alanine, to an initial OD600 ca. 0.01. The small inoculum was to mimic the conditions previously used in other *alr* deletion mutant study [11]. OD_{600} values were taken at 48h after inoculation.

To test D-alanine requirements in minimal media, *M. smegmatis* mc^2155, TAM23 and GPM292 were grown to an OD_{600} c.a. 1.0 in MADC supplemented with 50 mM D-alanine, washed twice with PBS-Tween and inoculated into pre-warmed minimal media (5 mM ammonium chloride, 22 mM dibasic potassium phosphate, 16 mM monobasic potassium phosphate, 25 pM ferric chloride, 8.6 nM zinc sulfate, 0.84 nM cobalt

chloride, 10.1 nM manganese chloride, 68 nM calcium chloride, 2.4 mM magnesium sulfate, 4.9 nM pyridoxal hydrochloride, 21mM glycerol, and 0.05 % Tween 80) at an initial concentration of approximately 1.2 x 10^6 c.f.u. /ml. Cultures were grown at 37^{O}C with shaking at 200 rpm in an Innova 4300 rotary incubator (New Brunswick Scientific Co.) and optical densities (OD_{600nm}) were monitored for approximately 200 h.

Generation and lag times were determined following described conceptual definitions [24]. To calculate generation times, logOD for each independent growth curve was plotted versus time during exponential growth. Linear regression lines were determined and from the slopes, generation times were calculated using the formula: G (min) = [1/(3.33 * slope)] * 60. Mean generation times with SEMs (n=3) were obtained from independent growth curves for each strain and condition. For growth curves displaying a plateau followed by resumption of exponential growth resembling diauxic growth, additional generation times were determined for each exponential growth phase.

Curves were compared by fitting individual nonlinear growth curves to each of the samples, followed by a MANOVA of the estimated growth curve parameters. Data analysis was carried out using the NLIN and GLM procedures (SAS Institute Inc. 2009. SAS OnlineDoc® 9.2. Cary, NC: SAS Institute Inc.). Pair-wise comparisons of generation times were performed using the Mixed Procedure in SAS version 9.1.3 (SAS Institute). *P* values were determined using least squares means and applying the Tukey adjustment. The two lag times were compared using the two-tailed unpaired *t*-test SAS procedure. Graphs were constructed using DeltaGraph version 5.6.2 (Red Rock Software).

9.2.5 Growth of M. smegmatis alr mutants in solid media.

To determine the effect of D-alanine on growth in solid media, *M. smegmatis* strains were streaked on agar media with and without 5.0 mM D, L-alanine (LBT), or with and without 50 mM D-alanine (MADC). A 1 μL disposable loop was used to streak cells taken from 50% glycerol stocks prepared from cultures grown in original stocking media (MADC for *M. smegmatis* wild type mc^2155 and TAM23(pTAMU3); MADC with 50 mM D-alanine for insertion mutant TAM23 and LBT with 5.0 mM D,L alanine for deletion mutant GPM292 and its complemented derivative GPM345. Plates were incubated for seven days at 37ºC and then imaged.

To determine the overall effect of solid media components on growth, glycerol stocks of *M. smegmatis* wild type and *M. smegmatis* TAM23 and GPM292 *alr* mutants were streaked on LBT, MADC and MADC agar plates supplemented with: 1mg/mL tryptone (Difco); a mixture of 20 aminoacids (0.25 mM each); a mixture of glucogenic aminoacids including L-alanine, L-asparagine, L-aspartate, L-glutamic acid and L-glutamine (1mM each); a mixture of branched chain aminoacids including L-isoleucine, L-leucine and L-valine (1.67 mM each); Middlebrook 7H10 and Midddlebrook 7H11. Growth patterns were determined after incubating for 7 days at 37ºC. Further testing was carried out streaking glycerol stocks of wild type mc^2155, complemented strain GPM345, and *alr* mutants TAM23 and GPM292 on MADC agar with or without D-alanine, D-glutamate, L-glutamate, D-serine or L-serine (5.0 mM each), or a mixture of 5.0 mM D-glutamate with 18 mM pyruvate. Plates were incubated for 7 days at 37ºC, imaged and scored. Scores were generated for growth coverage and colony size, as shown in Figure 9.1.

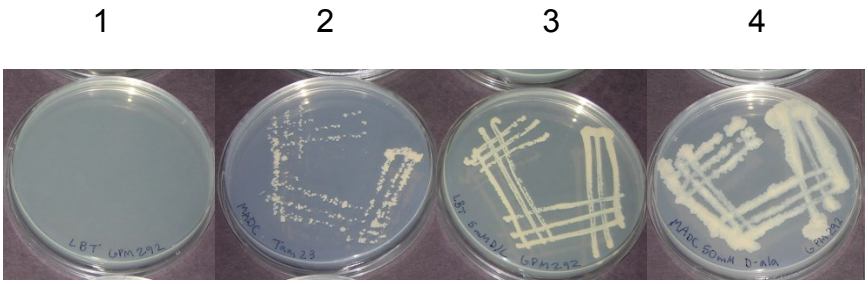


Figure 9.1. Scoring growth of *M. smegmatis* strains on agar media**.** To compare growth on solid agar media, the following scoring scheme was used, based on the following models: 1) GPM292 on LBT, -; 2) TAM23 on MADC,+; 3) GPM292 on LBT with 5.0 mM D/L-alanine,++; and 4) GPM292 on MADC 50 mM D-alanine, +++.

9.2.6 NMR Data Collection.

The NMR spectra were collected on a Bruker 500 MHz Avance spectrometer equipped with a triple-resonance, Z-axis gradient probe. A BACS-120 sample changer with Bruker Icon software was used to automate the NMR data collection. For the comparison between wild-type cells and the racemase mutants without D-alanine supplementation, one-dimensional (1D) ^{1}H NMR spectra were collected with a standard solvent presaturation sequence and a spectral width of 5482.5 Hz and 32k data points at 298K. A total of 16 dummy scans and 128 scans were used. For the comparison between wild-type cells and the racemase mutants supplemented with 50 mM D-alanine, 1D ^{1}H NMR spectra were collected using excitation sculpting to remove the solvent [25] with a sweep width of 5482.5 Hz and 32k data points at 298K. A total of 16 dummy scans and 128 scans were used.

The series of three time-zero two-dimensional (2D) ^{1}H-^{13}C HSQC spectra ($HSQC_0$) were collected using the pulse sequence described by Hu *et al.* [26]. The spectra were collected with 2048 data points and a spectral width of 5000.0 Hz in the ^{1}H dimension, and 64 data points and a spectral width of 17605.6 Hz in the ^{13}C dimension. A total of 16 dummy scans and 64 scans with a relaxation delay of 1.5 s were used to obtain each 2D ^{1}H-^{13}C $HSQC_0$ spectrum.

9.2.7 1D NMR Data Analysis.

All 1D ^{1}H NMR spectra were processed as previously described [27]. Data were processed automatically using ACD/1D NMR manager version 12.0 (Advanced

Chemistry Development, Inc). Intelligent bucketing was used to integrate each region with a bin size of 0.025 ppm and a looseness of 50%. The table of integrals was centered average and imported into SIMCA version 12.0 (Umetrics) for principal component analysis (PCA) or orthogonal partial least-squares discriminant analysis (OPLS-DA). Supervised classification for each group was arranged by cell type, with *M. smegmatis* mc^2155 defined as the control (or assigned a value of 0), and *M. smegmatis* TAM23 and GPM292 defined as the mutant strains (or assigned a value of 1). Ellipses for each group or class in the PCA or OPLS-DA scores plot depict the 95% confidence level from a normal distribution.

9.2.8 2D NMR Data Analysis.

2D ^{1}H-^{13}C HSQC spectra were processed using the NMRPipe software package [28]. The spectra were Fourier transformed, manually phased, and baseline corrected. The processed 2D ^{1}H-^{13}C HSQC spectra were then analyzed using NMRView [29] to assign chemical shifts, intensities, and volumes to each peak. Chemical shift lists were assigned to specific metabolites using the Human Metabalome Database [30], Madison Metabolomics Database [31], and the Platform for Riken Metabolomics [32]. A chemical shift error-tolerance of 0.05 ppm and 0.40 ppm was used for ^{1}H and ^{13}C chemical shifts, respectively. The identification of metabolites and metabolomics pathways from the NMR metabolomics data were further verified using the Kyoto Encyclopedia of Genes and Genomes (KEGG) [33] and MetaCyc [34] databases.

Metabolomic peak volumes from the $HSQC_0$ experiment was calculated as previously described [26, 35]. The calibration of metabolite peak volumes from the $HSQC_0$ experiments with an absolute concentration was performed using a series of 5 external standards, each consisting of 9 ^{13}C-labeled compounds with concentrations ranging from 5 μM to 300 μM. All standard samples also included 500 μM of TMSP as an internal standard for spectral referencing and normalization. A linear regression line was created to correlate the absolute concentrations with the average time-zero peak volume for each metabolite standard in order to create a standard calibration curve. The *M. smegmatis* mc2155, TAM23 and GPM292 metabolite concentrations were then calculated from the time-zero peak volumes obtained from the $HSQC_0$ experiment and from this standard calibration curve. The $HSQC_0$ experiment was obtained for a triplicate set of cultures for each *M. smegmatis* strain. The triplicate $HSQC_0$ experiments were used to calculate averages and standard deviations for each metabolite concentration.

9.2.9 Construction of TAM23 mutant library and isolation of D-alanine Dependent Mutants.

Transposon mutagenesis was carried out as described [36]. 5 mL of MADC broth supplemented with 50 mM D-alanine and kanamycin (20 μg/mL) were inoculated with 50 μL of a frozen stock of *M. smegmatis* TAM23. The culture was grown to an optical density OD_{620nm} of c.a. 1.0 and centrifuged at room temperature for 10 minutes at 3,700 x *g* (Sorvall Legend™ RT Kendro Laboratory products. Thermo Fisher Scientific Inc. Waltham, MA. USA). The pellet was washed twice in an equal volume of MP buffer and

resuspended in 5 mL of MP. This cell suspension was incubated at 37°C, standing, for 30 minutes before phage infection, and 100 µL were used to make ten-fold dilutions (10^{-1} - 10^{-6}) in MADC. Titration was carried out plating in MADC agar supplemented with 50 mM D-alanine and 20 µg/mL kanamycin. For transfection, 2.0 mL of undiluted *M. smegmatis* TAM23 were transferred to 50-mL screw cap tubes and mixed with 2.0 mL of a high titer lysate of phAE87 (Tn*5371*) (Table 9.1), ($1x10^9$ pfu/mL), to achieve a multiplicity of infection of 10:1. The mixture was incubated at 37°C, standing for 3 hours. Two milliliters of *M. smegmatis* TAM23 were incubated with the same quantity of MP buffer as negative controls. After incubation, cells were centrifuged for 10 minutes at 3,700 x *g* at room temperature (Sorvall Legend™ RT Kendro Laboratory products. Thermo Fisher Scientific Inc. Waltham, MA. USA). The pellet was resuspended in 2.0 mL of MADC broth and plated in MADC agar supplemented with 50mM D-alanine, 20 µg/mL kanamycin and 75 µg/mL hygromycin. Negative controls were plated in MADC agar with and without antibiotics. Plates were incubated at 37°C for 48 - 72 hours and individual transductants were isolated.

Individual transductant colonies were checked for D-alanine dependence through a two-step process. For the first step, each isolated colony was replica plated in MADC agar media with 20 µg/mL kanamycin and 75 µg/mL hygromycin in the presence or absence of 50 mM D-alanine. Wild type *M. smegmatis* mc^2155 and *M. smegmatis* TAM23 were used as controls in each inoculated plate. Plates were incubated for 7 days at 37°C before being replica plated once more in identical manner, to reconfirm D-alanine auxotrophy. An individual auxotroph (TAM23-12) was further characterized.

Table 9.2. Primers used for gene cloning experiments

Primer	Sequence (5'-3')
AlrKO1	AAATTTGGATCCGACAAGAACATGATCGACAA
AlrKO2	AAATTTGGTACCGGTCTGCATCGTCATAATCT
AlrKO3	AAATTTAAGCTTGGCAGGACAACAAGATTGAG
AlrKO4	AAATTTTCTAGAGCCCTCGACATCCATTGCTT
AlrInt1F	ATGACCTTGAAATGCCCGGTGG
AlrInt1R	ACCACGAACTGGTCCATGCAGAT
AlrIntF	CGGTCTGATGAGCCACCT
AlrIntR	CCTTGATCAGCACGTCGATA
5alr2	AAATTTCCATGGTGGGGCAGTACTACAACTTC
3alr2	AAATTTAAGCTTGTAGGGGTCTTCCTTGCTCA
Hygro-2F	AACACCTCGAAGTCGTGCAGGAAG
Hygro-2R	CTACCTGGTGATGAGCCGGATGAC
Tn-Mar-Hyg-1	AGTGCCACCTAAATTGTAAGCGT
RS6-4	GTAATACGACTCACTATAGGGCNNNNCATG
Tn-Mar-Hyg-2	GACCGAGATAGGGTTGAGTGTTGT
T7	TAATACGACTCACTATAGGG
Tn-MAR-Hyg-2b	CGAGATAGGGTTGAGTGTT
MURISMEGF	CCGGCCATATCTGACCTCATG
MURISMEGR	TGGCGACACAGATCATCT
TBALRUF	ACGGTACCGGCCTCATCCCTATTCGAC
TBALRUR	CGCTCTAGATGGCTTTCCGACATTCTCC
TBALRDF	GGCAAGCTTGAGGACGTATCACCAGGAC
TBALRDR	ATGCTCGAGCATGACGCCGCACGATAC
EXTBALRF	GTTAACGTGAAACGGTTCTGG
TBALRR	GGATACCCTCACGGCTCA

9.2.10 Identification of the Himar1 transposon insertion site and target gene in M. smegmatis TAM23-12.

M. smegmatis TAM23-12 cells (approx. 1-5 x 10^6 from a concentrated 50% glycerol stock) were directly lysed in a thermocycler (Eppendorf Mastercycler Gradient, Eppendorf, Hauppauge, NY). Samples were subjected to a nested PCR method adapted from published procedures [37, 38] using two rounds of amplification. The first amplification round was carried out with an annealing temperature of 52°C, using the transposon specific primer Tn-Mar-Hyg-1 and the degenerate primer RS6-4 (Table 9.2). Reactions were carried out in a total volume of 50 μL with 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 5% (v/v) dimethylsulfoxide (DMSO), 2 mM $MgCl_2$, 0.2 mM deoxynucleoside triphosphates, 0.1 μM Tn-Mar-Hyg-1, 1.0 μM RS6-4, and 2.5 units of Promega GoTAQ DNA polymerase (Promega Corporation. Madison, WI). For the second round, 1 μL of the first round PCR mixture was amplified using primers Tn-Mar-Hyg-2 and T7 (Table 9.2) under the same conditions as for the first round, except that primers were used at 0.5 μM. The mixture from the second nested PCR reaction was purified and concentrated using a batch column purification method (Promega Wizard PCR Preps) and run onto a 1.5% agarose gel. DNA bands were excised and purified using a Qbiogene Geneclean III kit (MP Biomedicals, LLC. Aurora, OH) and corresponding DNA sequences were determined at the University of Nebraska-Lincoln sequencing facility by the Sanger method using the sequencing primer Tn-MAR-Hyg-2b (Table 9.2). The information from this sequence was combined with the *himar1* transposon sequence and the Blastn suite

(http://blast.ncbi.nlm.nih.gov/) analysis to determine the exact location of the transposon insertion site within the target gene.

To confirm the identity of the target gene, PCR was carried out using primers specific to the target genes outflanking each transposon insertion site. Cells were lysed as described above. The *Himar1* insertion within the gene of interest should yield in the mutant strain a fragment 2,217 bp longer than in the wild type strain mc^2155. Amplification was carried out using a standard procedure in a 50 μL reaction mixture containing 25 ng of each specific primer pair in the presence of 1x NH_4-based Reaction Buffer (Bioline USA Inc., Taunton, MA), 2 mM $MgCl_2$, 5% (v/v) DMSO, 0.3 mM deoxynucleoside triphosphates, and 2.5 units of Biolase™ DNA Polymerase (Bioline). *M. smegmatis* genomic DNA was amplified using primers MURISMEGF and MURISMEGR (Table 9.2), spanning the region external to the identified target gene. The reaction was carried out with thermocycler settings of 94°C for 5 minutes, followed by 30 cycles of 94°C for 1 min, 50°C for 30 sec, 72°C for 1 min, and a final extension at 72°C for 7 min.

9.2.11 Complementation and phenotypic characterization of M. smegmatis D-alanine auxotrophic strain TAM23-12.

To construct the complementing vector pBUN373, the tetracycline resistance gene from pBR322 [39] was released by digestion with *Bsg*I and *Eco*RI and ligated into the *Hpa*I and *Eco*RI sites of the *E. coli-Mycobacterium* multi-copy shuttle plasmid pMV203 [16]. The corresponding recombinant clone (pBUN73) was digested with *Spe*I

and then partially digested with *Nhe*I to remove the kanamycin resistance gene, resulting in the 4,496 bp tetracycline resistant shuttle plasmid pBUN369. A 2,361 bp fragment containing the *M. smegmatis alr* gene was released from pBUN82 [16] by *Eco*RI and *Acc*65I digestion and the final step in pBUN373 construction was to ligate the fragment into *Eco*RI - *Acc*65I digested pBUN369 (Table 9.1). After transformation with pBUN373 the recombinants were incubated at 37°C for 2-3 days on MADC agar supplemented with 50 mM D-alanine and 12.5 µg/mL tetracycline as a mycobacterial selection marker [40]. Non-transformed *M. smegmatis* parent strains were used as controls. Tetracycline resistance was confirmed by culturing individual colonies in 5.0 mL of MADC broth supplemented with 50 mM D-alanine and 12.5 µg/mL tetracycline at 37°C in a shaker at 250 r.p.m. (Unimax 1010. Heidolph Instruments GmbH&Co. KG. Schwabach. Germany) until the culture reached an OD_{620nm} of c.a. 1.0. The cultures were mixed with seven 4 mm glass beads (Paul Marienfeld GmbH & Co. KG, Am) before the optical densities were measured. Cells were concentrated by centrifugation at room temperature for 10 minutes at 3,700 x *g* (Sorvall Legend^{TM} RT Kendro Laboratory products. Thermo Fisher Scientific Inc. Waltham, MA. USA). The supernatant was discarded and the pellet was washed twice in equal volume of MADC. The final pellet was resuspended in 4 mL of MADC and the OD_{600} was measured again in order to inoculate the same amount of cells in each culture. 100 µL of dilutions 10^{-5} and 10^{-6} were inoculated by triplicate in MADC agar supplemented with 50 mM D-alanine and 12,5 µg/mL tetracycline and in MADC agar supplemented only with 12,5 µg/mL tetracycline. Plates were incubated at 37°C for 7 days and examined visually at day 3, 5 and 7.

Phenotypic characterization included growth in media with and without D-alanine, as well as macro and microscopic visualization of colonies. Strains were grown to an $O.D._{600nm}$ c.a. 1.0 in MADC broth (*M. smegmatis* mc^2155), MADC broth supplemented with 50 mM D-alanine and 10 µg/mL kanamycin (*M. smegmatis* TAM23), and MADC broth supplemented with 50 mM D-alanine, 10 µg/mL kanamycin and 75 µg/mL hygromycin (*M. smegmatis* TAM23-12). Strains were sub-cultured in 5.0 mL of the same media as above and incubated at 37°C in shaker at 250 r.p.m. (Unimax 1010. Heidolph Instruments GmbH&Co. KG. Schwabach. Germany) until the culture reached an OD_{600nm} c.a. 1.0. The cultures were mixed using seven 4 mm glass beads (Paul Marienfeld GmbH & Co. KG, Am) before the OD_{600nm} was measured. Cells were concentrated by centrifugation at room temperature for 10 minutes at 3,700 x *g* (Sorvall Legend^{TM} RT Kendro Laboratory Products, Thermo Fisher Scientific Inc., Waltham, MA. USA). The supernatant was discarded and the pellet washed twice in equal volume of MADC. The final pellet was resuspended in 4.0 ml of MADC and the OD_{600nm} was measured again in order to inoculate the same amount of cells in each culture. 100 µL of dilutions 10^{-5} and 10^{-6} were inoculated by triplicate in the follow media: MADC agar, MADC agar supplemented with D-alanine, MADC agar supplemented with kanamycin, MADC agar supplemented with D-alanine and kanamycin, MADC agar supplemented with hygromycin, MADC agar supplemented with D-alanine and hygromycin, MADC agar supplemented with kanamycin and hygromycin and MADC agar supplemented with D-alanine, kanamycin and hygromycin as above. Plates were incubated at 37°C for 7 days and examined visually and microscopically at days 3, 5 and 7.

9.2.12 Vectors for M. tuberculosis alr inactivation and merodiploid constructions.

To construct the vector for chromosomal *alr* inactivation, *M. tuberculosis* H37Rv genomic DNA was amplified with primer pairs TBALRUF / TBALRUR and TBALRDF / TBALRDR (Table 9.2). The purified 613 bp PCR product, corresponding to the upstream region and 33 bp at the 5' end of the *alr* gene, was digested with *Kpn*I and *Xba*I. The fragment was cloned into pYUB854 digested with the same enzymes, generating plasmid pBUN287. The purified 625 bp PCR product, corresponding to the downstream region and 43 bp at the 3' end of the *alr* gene, was digested with *Hin*dIII and *Xho*I and ligated to pBUN287 digested with the same enzymes. The resulting plasmid, pBUN288, was digested with *Pac*I and ligated with phAE87 digested with the same enzyme. The resulting phasmid phBUN1, carrying the upstream and downstream regions of the *M. tuberculosis alr* gene flanking a hygromycin resistant marker, was used to inactivate the chromosomal copy of the *alr* gene in *M. tuberculosis*, as described [22].

For the construction of recombinant plasmid pBUN307, the complete coding region of *alr* from *M. tuberculosis* H37Rv was amplified using the Expand High Fidelity PCR System (Roche). The 1,227 bp *M. tuberculosis alr* was amplified from genomic DNA as a 1248 bp fragment using primers EXTBALRF and TBALRR (Table 9.2). After purification with the Wizard PCR Prep Kit (Promega), the PCR fragment was cloned into the Original TA Cloning Kit vector pCR2.1 (Invitrogen). The corresponding recombinant clone (pBUN201) was digested with *Mse*I and *Bam*HI and ligated into the *Nde*I-*Bam*HI sites of the vector pET15b (Novagen). The recombinant clone (pBUN205) was verified by restriction endonuclease digestion and DNA sequencing. For the construction of

pBUN307, pBUN205 was linearized with *Xba*I and polished with Pfu polymerase. This blunt ended linear fragment containing the H37Rv *alr* was further digested with *Eco*RI and ligated into the *Pvu*II and *Eco*RI sites of the *E. coli-Mycobacterium* multi-copy shuttle plasmid pMV203 [16]. In this construct, major recombinant plasmid transcripts are expressed in the same direction.

9.2.13 Construction, verification and characterization of M. tuberculosis alr mutant strains.

Standard procedures were used to transform wild type *M. tuberculosis* H37Rv with pBUN307, the multicopy plasmid carrying the *M. tuberculosis* wild type *alr* gene [22]. The resulting merodiploid strain H37Rv (pBUN307) was selected in media with 10 μg/mL kanamycin and the presence of the extrachromosmal *alr* copies was verified by PCR. Plasmid DNA isolation was used to confirm the generation of a *M. tuberculosis* H37Rv recombinant carrying the plasmid pBUN307. Inactivation of the *alr* gene in wild type and merodiploid *M. tuberculosis* strains was carried out by transduction with phBUN1 at M.O.I. 1:10, following standard procedures [22]. Transfectants were selected by incubating for 6 weeks at 37°C in the presence of 50, 125 or 225 mM D-alanine (Sigma-Aldrich, CO. St. Louis, MO, USA) and 75 μg/mL hygromycin (Alexis Biochemicals, San Diego, CA, USA).

Inactivation of the chromosomal *alr* gene was confirmed by Southern blotting analysis. To this end, approximately 3.0 μg of *Sac*II (New England Biolabs, Ipswich, MA) digested chromosomal *M. tuberculosis* H37Rv wild type and mutant genomic DNA

was transferred to a nylon membrane (Hybond-N+. Amersham Pharmacia Biotech, Buckinghamshire, England UK), prehibridized and hybridized at 56 ºC, following standard procedures [23, 41]. The 0.6 kb PCR fragment generated by pBUN288 DNA amplification with primers TBALRDF and TBALRDR (Table 9.2) was used as hybridization probe. The fragment was radio labeled with [^{32}P] dCTP (Easy Tides 32P dCTP 3000 Ci/mMol; Waltham, MA) using a commercial kit (Rediprime II Kit, Amersham Pharmacia Biotech, Buckinghamshire, England), following the manufacturer's instructions. Washes were done under high-stringency conditions at 65ºC. The membrane was exposed for two hours at room temperature on a phosphor screen (Bio-Rad Laboratories; Hercules, CA). Radioactivity detection was carried out in a Molecular Imager PharosFX Plus system (BioRad Laboratories; Hercules, CA), following manufacturer's instructions.

Phenotypic studies included determination of acid-alcohol resistance by the Kinyoun stain [42] and, macro and microscopic visualization of colonies (Nikon Instruments, Inc.). Wild type *M. tuberculosis* H37Rv, merodiploid *M. tuberculosis* H37Rv pBUN307 and *M. tuberculosis* H37Rv TBcol-1 *alr* mutant were initially grown in MOADC broth, with or without supplements, to an $OD_{620\,nm}$ of approximately 1.0. 10 µL (approximately $3x10^6$ cells) of each culture were inoculated in 7 mL MGIT® tubes (Becton, Dickinson and Company, Sparks, MD, USA) supplemented with OADC (for wild type *M. tuberculosis* H37Rv), OADC and kanamycin 10 µg/mL (for recombinant *M. tuberculosis* H37Rv pBUN307 strains) and OADC with or without 50 mM D-alanine and hygromycin 75 µg/mL (for *M. tuberculosis* H37Rv TBcol-1 *alr* mutant strain). The tubes

were incubated in MGIT® 960 System (Becton, Dickinson and Company, Sparks, MD) at 37°C until mycobacterial growth was detected. Positive tubes were centrifuged for 15 minutes at 4000 r.p.m. at 4°C, 10 µL of sediment were plated in MADC agar plates with or without supplements, and incubated for 3 weeks at 37°C. Macro and microscopic examination were carried out weekly in these plates. Additionally, 20 µL of sediment were used to determine acid-alcohol resistance following standard methods [42].

9.3 Results

9.3.1 Construction, verification and complementation of M. smegmatis alr deletion mutant.

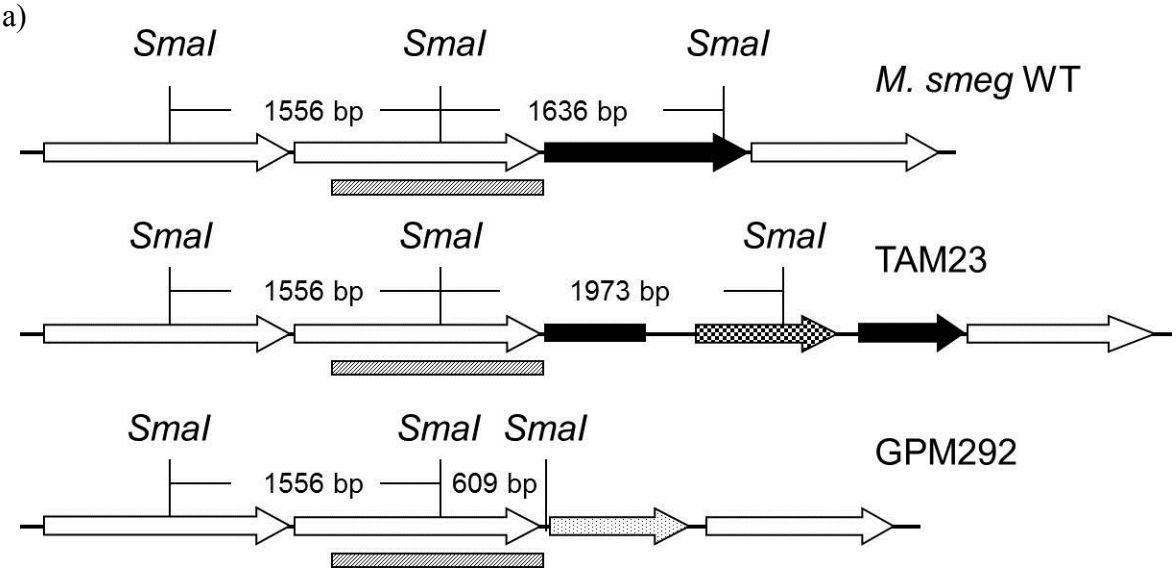
To solve the controversy surrounding the requirement for the *M. smegmatis alr* gene in the absence of D-alanine, we generated *alr* deletion mutants GPM292, GPM293, GPM294 and GPM295. After conducting a preliminary characterization by Southern blot analysis (see below), one representative deletion mutant, *M. smegmatis* GPM292, was selected for further analysis. Deletion of 99% of the *alr* gene coding sequence was accomplished using an approach that was similar to that previously described for the construction of other *M. smegmatis alr* deletion mutant strains [11]. To that end, we transformed the wild type strain mc^2155 with a pBUN381 (Table 9.2) fragment carrying the non-polar pUC18K *aphA-3* kanamycin resistance marker outflanked by the *M. smegmatis alr* upstream and downstream regions. The inactivation was accomplished using a one-step strategy, since the lack of illegitimate recombination in this species makes this direct selection for a double cross-over feasible [40].

Our *alr* deletion mutants are in principle identical, except for polylinker site sequences introduced at the drug marker region, to the *alr* deletion mutants previously described, including the same non-polar *aphA-3* marker [11]. A new deletion mutant was made to assure that both the deletion and insertion mutants to be analyzed were derived from the same parental strain, a low-passage isolate of mc^2155 that was directly received from the original source [15]. Thus, any potential source of variability due to mc^2155 isolates kept at different laboratories was eliminated for the comparison of insertion and deletion *alr* mutant strains. Mutant strains were verified by PCR (data not shown) and Southern blotting analysis. The schematic display of *M. smegmatis* wild type and *alr* mutant patterns in Southern blotting analysis with a left flank probe upon *Sma*I digestion predicts the following approximate sizes for the band patterns: 1.63 kb and 1.56 kb for wild type mc2155; 1.97 kb and 1.6 kb for TAM23 and 1.56 kb and 0.6 kb for the deletion mutants (Figure 9.2a). *M. smegmatis* GPM 292 - 295 yielded the expected patterns and were confirmed to be true *alr* deletion mutants (Figure 9.2b). The complemented strain GPM345 was generated by transformation of *M. smegmatis* GPM292 with pTAMU3 (Table 9.1) and verified by the presence of a 0.56 kb signature band corresponding to the amplification of the pTAMU3 hygromycin resistance gene (data not shown).

9.3.2 Enzymatic activity of M. smegmatis alr mutants.

Alr enzymatic activity was determined using a 1-step chiral plate assay, designed to directly measure the conversion of ^{14}C-D-alanine into ^{14}C-L-alanine without the need for further coupling steps that could diminish the sensitivity of the assay (Figure 9.3). The

a)


SmaI
SmaI
SmaI
M. smeg WT
1556 bp
1636 bp
SmaI
SmaI
SmaI
TAM23
1556 bp
1973 bp
SmaI
SmaI
SmaI
GPM292
1556 bp
609 bp

b)

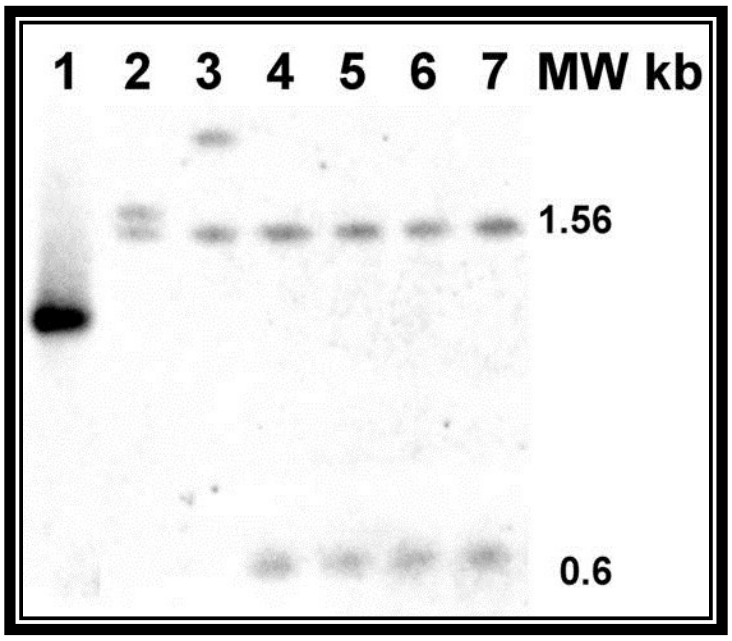

1
2
3
4
5
6
7
MW kb
1.56
0.6

Figure 9.2. a) Genomic comparison of *M. smegmatis* wild type and mutant *alr* loci. Location of left flank probe complementary sequences in relevant *SmaI* fragments for WT, TAM23, and GPM292 *alr* chromosomal loci. Displayed features are: *alr* upstream and downstream sequences adjacent to the original chromosomal locus (open arrows), *alr* gene (filled arrows), probe sequences (striped boxes), and the *kan* markers aph2 (checkered) and aphA-3 (dotted) arrows. **b)** Southern blotting analysis of *M. smegmatis* wild type *and alr* mutants. Genomic DNA was digested with *SmaI* (lanes 2 to 7). Lane 1 contains the non-labeled 1,237 bp probe homologous to the *alr* upstream flanking sequence defined by amplification with primers AlrKO1 and AlrKO2. Lane 2, mc2155; lane 3, TAM23; lanes 4-7, *alr* deletion mutants GPM292-GPM295.

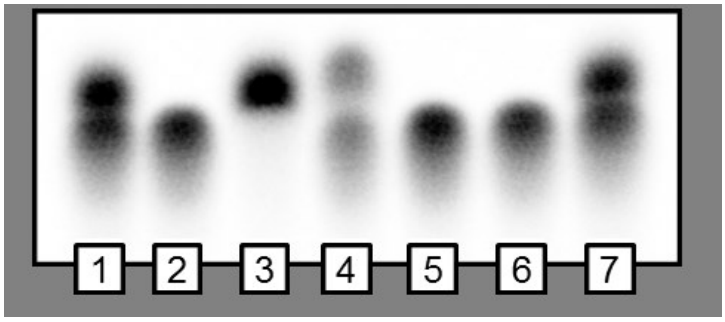


Figure 9.3. Alr enzymatic assay using chiral separation of alanine enantiomers. Protein extracts were incubated at 37ºC for 15 min in the presence of 50 mM Tris HCl pH 8.0, 15 mM D-alanine, 0.1 mM pyridoxal phosphate, and ^{14}C-D-alanine (lower spot). Control reactions were performed under the same conditions using WT-heat inactivated extracts with ^{14}C-L-alanine, ^{14}C-D-alanine or both radioactive alanine compounds present at a 1:1 ratio. Chiral plates were baked at 100ºC for 30 min, spotted with 5 µl of reaction mixture, air dried and developed in a saturated chamber with acetone, methanol and ddH_2O in a ratio of 10:2:2. **Lane 1,** 1:1 mixture of ^{14}C-D and ^{14}C-L-alanine; **Lane 2,** ^{14}C-D-alanine; **Lane 3,** ^{14}C-L-alanine; **Lane 4,** WT *alr*; **Lane 5,** TAM23; **Lane 6,** GPM292 & **Lane 7,** 1:1 mixture of ^{14}C-D and ^{14}C-L-alanine. Assays were performed in triplicate. Image was captured in BioRad Quantity One version 4.4.6, Contrast was adjusted uniformly and image cropped to remove excess space only.

test was able to resolve an equimolar mixture of D- and L-alanine (lanes 1 and 7), with each enantiomer generating a unique spot (lanes 2 and 3). In contrast to the wild type strain, able to convert D-alanine into an approximately equimolar mixture (44.4 ± 1.8 %) of both enantiomers upon 15 min incubation (lane 4), both *alr* insertion (lane 5) and deletion (lane 6) mutants were unable to convert D-alanine into L-alanine. Overnight incubation resulted in enantiomeric equimolar equilibrium (50.7 ± 1.6 %) in the wild type strain, as expected from the predicted thermodynamic parameters [43], but was unable to detect any enzymatic activity in the *alr* mutant strains (data not shown). Alr specific activity in wild type *M. smegmatis* was 0.2 μmol min^{-1} per μg of protein, consistent with previous data [10].

9.3.3 Growth of M. smegmatis alr mutants in different media.

The pair of isogenic insertion (TAM23) and deletion (GPM292) *alr* mutants was tested under various conditions. As expected, the composition of the media played a significant role in the ability of both mutant strains to grow with or without D-alanine. Minimal concentrations of D-alanine required for the mutant strains to yield significant growth were determined in duplicate using both LBT and MADC broth, with or without D-alanine supplementation. Upon incubation for 48 h., both *alr* mutants failed to grow in either media in the absence of D-alanine, supplementation with 0.1 mM D-alanine was sufficient for *alr* mutants to display some growth on MADC, and 0.3 mM D-alanine was required for growth in LBT (Figure 9.4). The results on LBT are basically identical to those previously reported for other *alr* deletion mutant [11]. In one experiment, CFUs

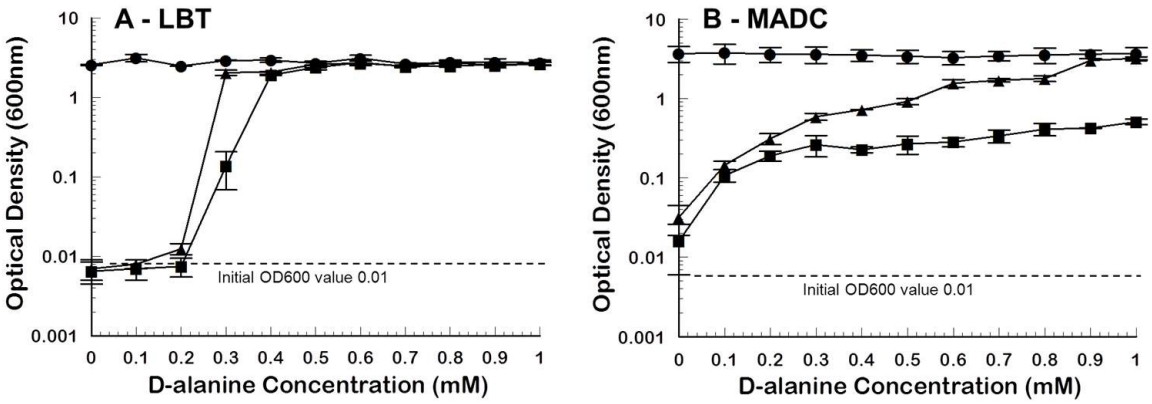


Figure 9.4. Growth of *M. smegmatis alr* mutants in MADC-Tween and LBT at varying D-alanine concentrations. *Mycobacterium smegmatis alr* deletion mutant GPM292 insertion mutant TAM23, and wild type mc^2155 were grown in both MADC-Tween **a)** and LBT **b)** media with varying concentrations of D-alanine. Cultures were inoculated to an initial OD600 ca. 0.01, and OD_{600} values were taken at 24h (not shown) and 48h (shown) after inoculation. Plotted data are means of two independent experiments (n=2). In one experiments, CFU were also determined and the results followed closely the data from the OD_{600nm} measurements depicted in this figure. Complemented strains TAM23 (pTAMU3) and GPM345 exhibited the same growth as mc^2155 (data not shown, n=1).

were also determined and the results followed closely the OD_{600nm} measurements. Complemented strains TAM23 pTAMU3 and GPM345 exhibited the same growth as mc^2155 (data not shown).

Growth curves of both *M. smegmatis alr* mutants TAM23 and GPM292 in minimal media with and without D-alanine exhibited similar patterns to those we reported before for TAM23 [10], including a long lag time in the absence of D-alanine. The wild type growth pattern was restored by addition of 50 mM D-alanine. Statistical analysis indicated that each strain and condition resulted in a characteristic growth curve ($P < 0.03$ or lower), except for mc^2155 and TAM23 grown in the presence of D-alanine that yielded similar growth patterns ($P < 0.61$). Most importantly, in medium without D-alanine, both *alr* mutant strains were able to grow, but displayed a long lag time (Figure 9.5).Thus, both *M. smegmatis alr* insertion and deletion mutants are not dependent on D-alanine for growth.

Observations in solid media confirmed similar growth patterns for *M. smegmatis* insertion (TAM23) and deletion (GPM292) mutants. Growth of the wild type and mutant strains was initially assessed on MADC and LBT agar with and without D-alanine. *M. smegmatis* mc^2155 and the alr^+ complemented strains displayed abundant growth on MADC and LBT agar both with and without D-alanine. Both TAM23 and GPM292 *alr* mutants were able to grow in MADC in the presence (more growth) or absence (poor growth) of D-alanine, while they grew on LBT only if supplemented with D-alanine, as it had been reported for other *M. smegmatis alr* deletion mutants [11]. Thus, these experiments conclusively demonstrate that D-alanine auxotrophy is conditional on the

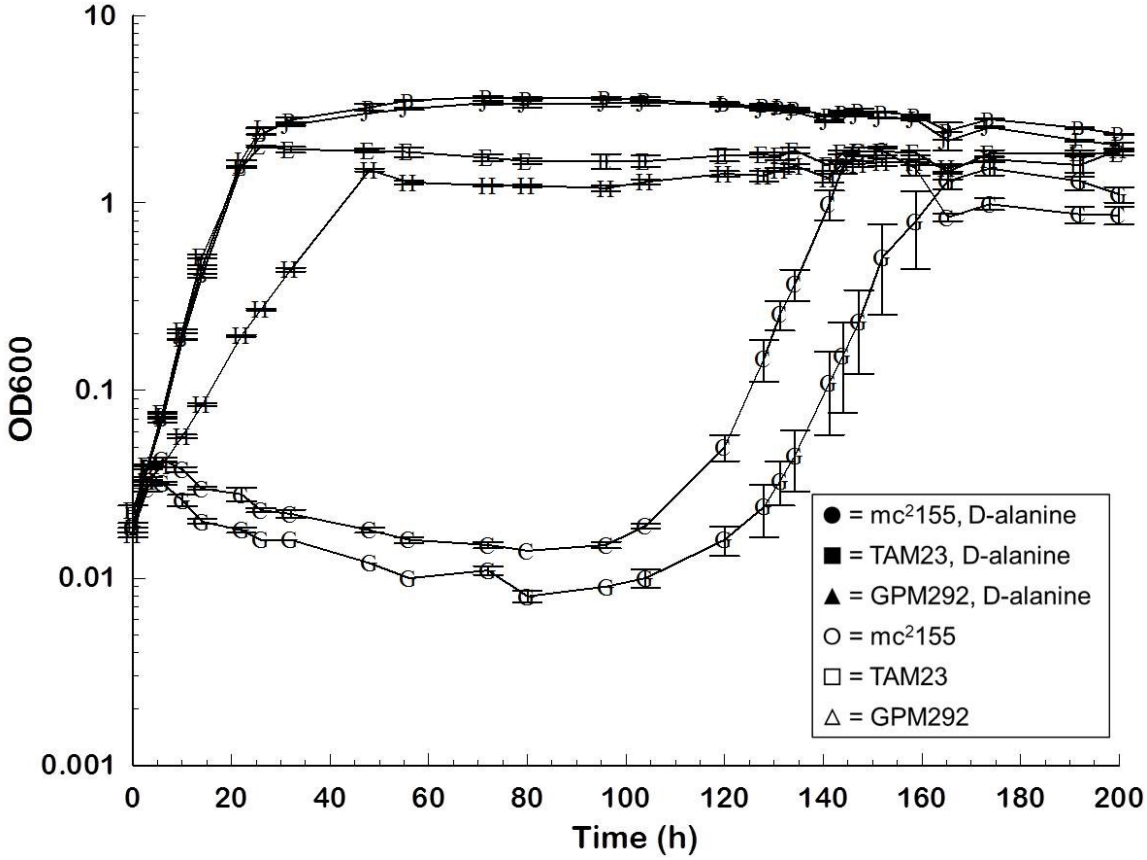


Fig. 9.5. Growth of *M. smegmatis* Alr mutant strains in minimal media with and without D-alanine. Strains mc^{2}155, TAM23, and GPM292 were grown in MADC-Tween with 50mM D-alanine to an OD_{600nm} of 0.8 – 1.2. Cells were harvested, washed twice with phosphate-buffered saline (PBS)–0.05% Tween 80, and three independent cultures per strain were inoculated 1:50 (initial OD_{600nm} ca. 0.02) into minimal media (Chacon et al. 2002) with 21 mM glycerol as the carbon source, 5.0 mM ammonium chloride, supplemented with or without 50 mM D-alanine. Growth was recorded for 200 h by monitoring the OD_{600nm}. SEM (n=3) are indicated by the corresponding error bars.

growth medium (Figure 9.6).

Plating on non-supplemented and supplemented LBT and MADC agar, revealed that the growth of *alr* mutants was inhibited by addition of Tryptone (LBT component absent in MADC), an equimolar mixture of 20 aminoacids or a mixture of glucogenic aminoacids (L-alanine, L-glutamic acid, L-glutamine, L-aspartate and L-asparagine). Mutant growth was also inhibited in Middlebrook 7H11 but not in Middlebrook 7H10 (Table 9.3). Further testing, using the plating score described in Figure 9.1, revealed abundant growth of *M. smegmatis* wild type mc^2155 and the complemented strain GPM345 in all media. *M. smegmatis* TAM23 and GPM292 exhibited wild type growth levels in the presence of 5 mM D-alanine, as well as in the presence of a combination of 5 mM D-glutamate and 18 mM pyruvate. Growth was observed with no D-alanine or any other supplementation, and with D-glutamate, L –glutamate, D-serine (Table 9.4), or 18 mM pyruvate (see Supplemental material). Addition of L-alanine, L-asparagine, L-serine or L-glutamine did not affect the growth of the wild type strain while totally inhibited the growth of both *alr* mutants (data not shown).

9.3.4 Global changes in the metabolome of alanine racemase mutants.

The principal component analysis (PCA) 2D scores plot (Figure 9.7a) illustrates a comparison between the metabolomes of *M. smegmatis* wild-type mc^2155 cells, and the alanine racemase insertion (TAM23) and deletion (GPM292) mutants grown in MADC without D-alanine. The 2D scores plot reduces the complex 1D ^1H NMR spectrum that

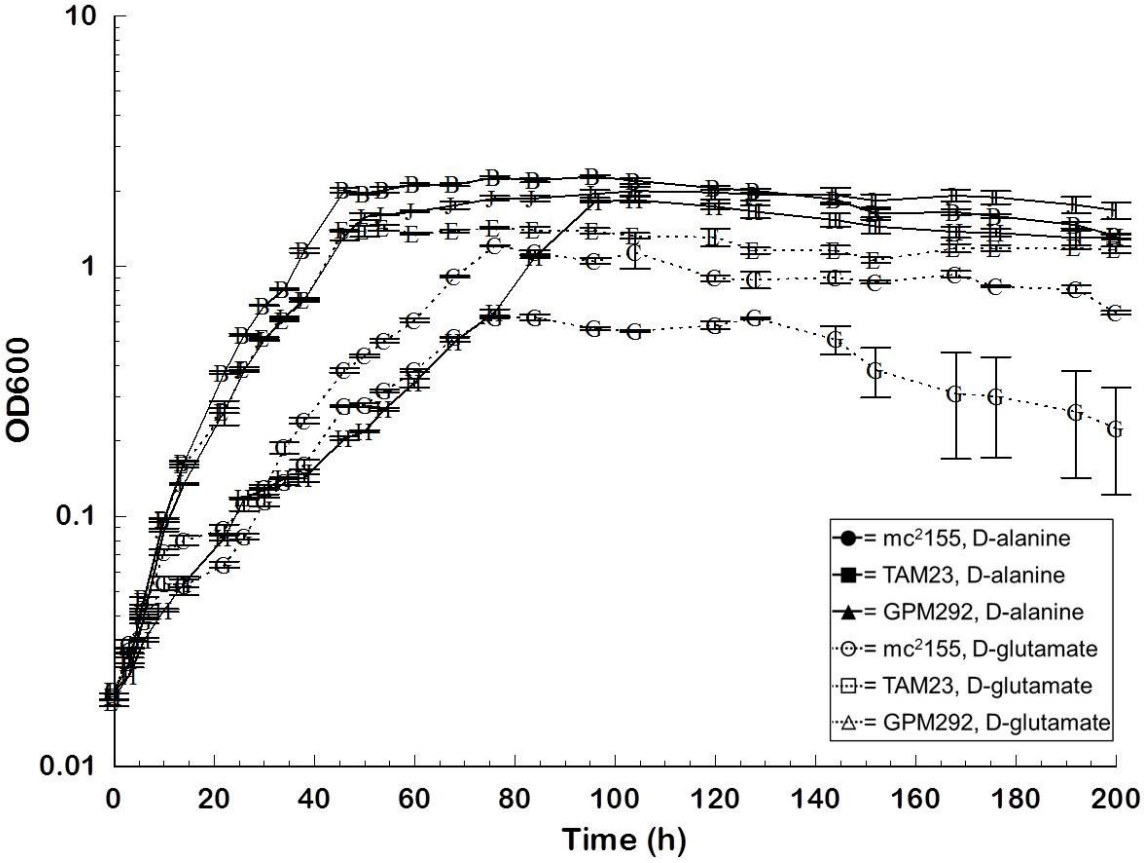


Fig. 9.6. Growth of *M. smegmatis* Alr mutant strains in minimal media with pyruvate as carbon source, supplemented with D-alanine or D-glutamate. Strains mc^2155, TAM23, and GPM292 were grown in MADC-Tween with 50 mM D-alanine to an OD_{600nm} of 0.8 – 1.2. Cells were harvested, washed twice with PBS–0.05% Tween-80, and inoculated 1:50 (initial OD_{600nm} ca. 0.02) into minimal media (Chacon et al. 2002) with 18 mM pyruvate replacing 21 mM glycerol as carbon source, 5.0 mM ammonium chloride, and 50 mM D-alanine (solid lines) or 5.0 mM D-glutamate (dotted lines). Growth was recorded for 200 h by monitoring the OD_{600nm}. SEM (n=3) are indicated by the corresponding error bars.

Table 9.3. Growth of *alr* mutants on selected agar media supplemented with various nutrients.

Media (Supplements)[a]	WT	TAM23	GPM292
LBT	+	-	-
MADC	+	+	+
+ Tryptone	+	-	-
+ 20 AA	+	-	-
+ Glucogenic AA	+	-	-
Middlebrook 7H10	+	+	+
Middlebrook 7H11	+	-	-

[a] MADC agar plates (regular and supplemented with 5mM of the listed amino acid) were streaked from glycerol stocks with 1 µl loops in four directions (S->E->N->W) on duplicate plates. Plates were incubated at 37 ºC and checked at day 7 for growth.). Results were recorded as growth (+) or no growth (-). Media tested were LBT, MADC, Middlebrook 7H10, Middlebrook 7H11, and MADC with and without supplementation of 1g/L Tryptone (Difco); mixture of 20 individual amino acids (0.25 mM each), and glucogenic aminoacids (1.0 mM each), as described in the text.

Table 9.4. Growth of *M. smegmatis* wild type and *alr* mutant strains on supplemented solid media.

Media (Supplements)[a]	mc2155	GPM292	TAM23	GPM345
MADC	+++	++	++	+++
MADC + 5.0 mM D-alanine	+++	+++	+++	+++
MADC + 5.0 mM D-glutamate	+++	++	++	+++
MADC + 5.0 mM D-serine	+++	++	++	+++
MADC + 5.0 mM D-glutamate + 18.1 mM pyruvate	+++	+++	+++	+++

[a] MADC agar plates were streaked from glycerol stocks (S->E->N->W) on duplicate plates. Plates were incubated at 37 ºC and checked after 5 and 7 days. Results reported are from day 7. Scores were assigned following the scheme described in Figure 9.1.

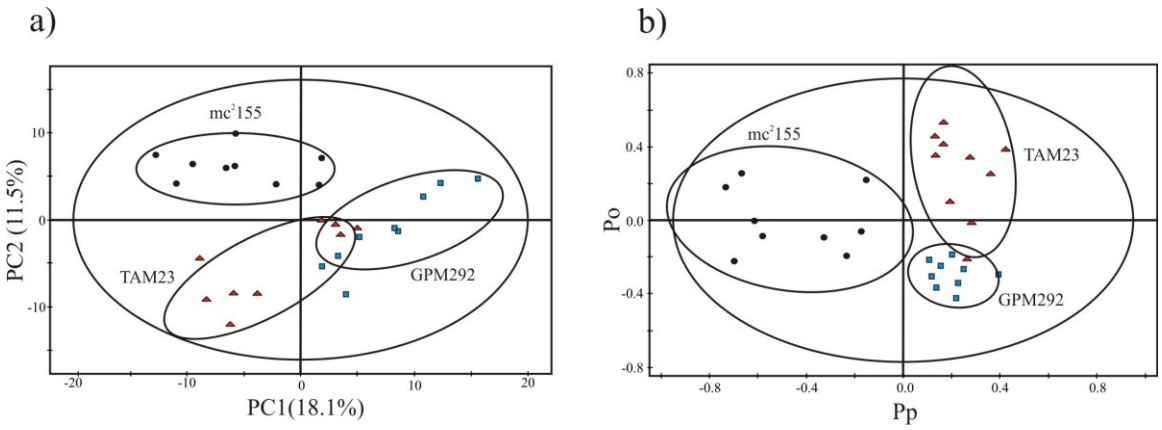


Figure 9.7. a) PCA 2D scores plot and **a)** OPLS-DA 2D scores plot. Comparison of the 1D ^{1}H NMR spectra obtained from *M. smegmatis* TAM23, GPM292, and mc^{2}155 cell lysates. The wild-type *M. smegmatis* mc^{2}155 cells were designated the control class in the OPLS-DA, and the remainder of the cells was designated as the mutant class. The OPLS-DA used one predictive component and one orthogonal component to yield a R^2X of 0.517, R^2Y of 0.822, and a Q^2 of 0.746. Cross validation using CV-Anova yielded a p-value of $1.38e^{-6}$. The ellipses correspond to the 95% confidence limits from a normal distribution for each cluster.

captures the state of the cellular metabolome into a single point in principal component space (PC1, PC2, etc.). Correspondingly, the principal components represent the largest variations between the individual 1D ^{1}H NMR spectra. The 2D scores plot indicates that PC1 and PC2 account for 18.1% and 11.5% of the variation in the NMR spectra,respectively. A total of 10 replicates for each of the mc^{2}155, TAM23, and GPM292 strains formed three distinct clusters in the PCA 2D scores plot. It is interesting to note, that despite the fact that both TAM23 and GPM292 lack alanine racemase activity, the two *M. smegmatis* strains formed separate clusters in the PCA 2D scores plot indicating different changes in the metabolome. More specifically, the GPM292 cluster was primarily separated from the mc^{2}155 and TAM23 clusters along PC1, the largest variation between the 1D ^{1}H NMR spectra.

The unsupervised PCA model effectively demonstrated class separation between the three *M. smegmatis* strains, which is generated based on the highest variations in the NMR data. To gather information about the discriminatory variations that maximize class separation, OPLS-DA was performed [44]. The OPLS-DA 2D scores plot exhibited a similar class separation between *M. smegmatis* mc^{2}155 wild-type cells and the two alanine racemase mutants (Figure 9.7b). The OPLS-DA model yielded a reliable fit as evident by an $R^{2}X$ of 0.517, $R^{2}Y$ of 0.822, and a Q^{2} of 0.746. Importantly, cross validation using CV-Anova [45] yielded a statistically significant p-value of $1.38e^{-6}$, validating the class separation between the two alanine racemase mutants and the mc^{2}155 wild-type cells. More importantly, the alanine racemase mutant strains clustered at different locations along the orthogonal component of the 2D scores plot, demonstrating

a statistically significant class separation between the deletion and insertion alanine racemase mutants. This is consistent with the PCA 2D scores plot where three distinct clusters were formed, indicating differences between the three *M. smegmatis* strains. Furthermore, the S-plot generated by OPLS-DA identified the NMR spectral differences and, correspondingly the major metabolomic changes contributing to the class separations (Figure 9.8). The significant variables (chemical shifts) labeled on the S-plot had contributions of 0.1 or greater (correlated) or -0.1 or less (anti-correlated). The metabolites identified from TAM23 and GPM292 as major contributors to class separation are summarized (Table 9.5). Specifically, a pair wise comparison between the alanine racemase insertion mutant (TAM23) and the wild-type mc^2155 cells identified an increase in TAM23 for serine, glycolate, phosphoserine, alanine, lysine, UDP, and ATP. Conversely, a decrease was observed for glutamate, isoleucine, lactate and TCA cycle intermediates in TAM23. The S-plot resulting from the pair wise comparison between the alanine racemase deletion mutant (GPM292) and the wild-type mc^2155 cells identified a larger number of bins significantly contributing to the class separation. However, as in TAM23, serine, glycolate, phosphoserine, lysine, alanine and UDP were increased in GPM292. Similarly, TCA cycle intermediates and glutamate were also decreased in GPM292. Despite the general similarity between TAM23 and GPM292, there were some differences in specific metabolites (Table 9.5). For example, acetate was increased in TAM23, but decreased in GPM292. Conversely, lactate increased in GPM292, but decreased in TAM23.

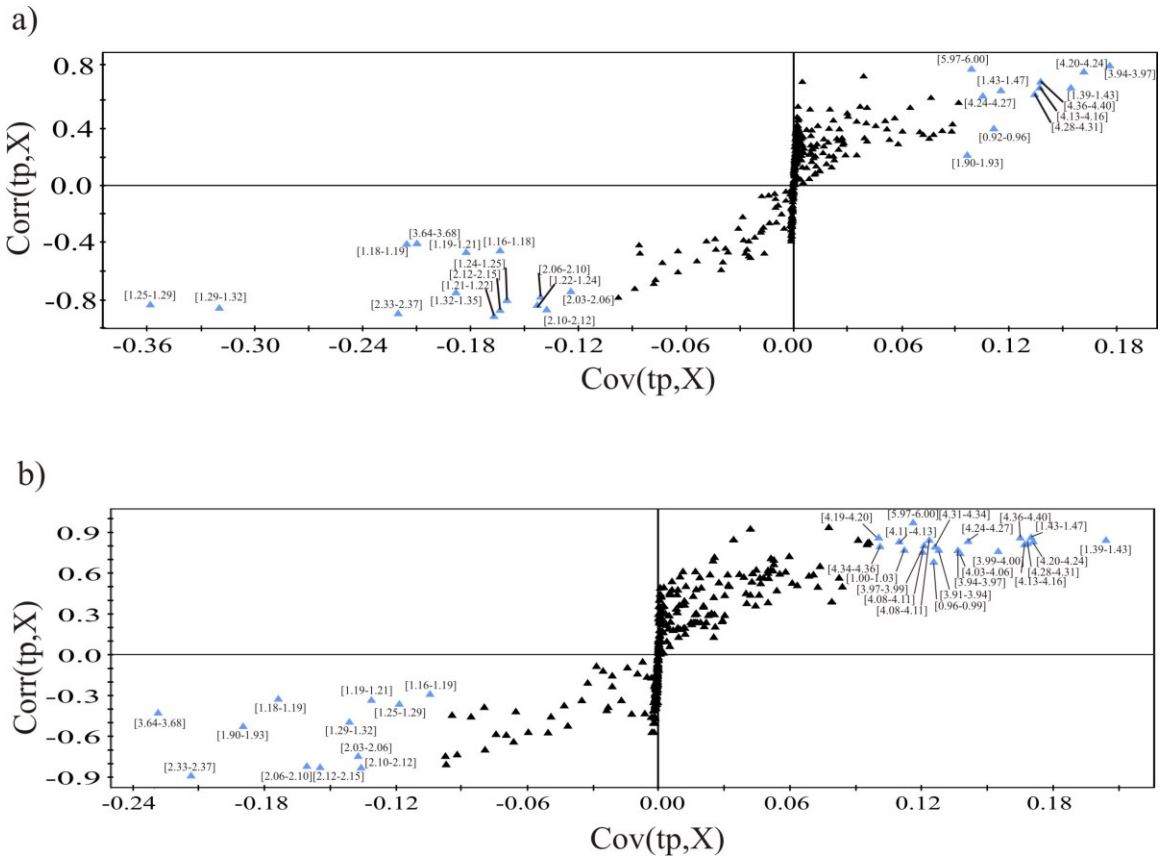


Figure 9.8. OPLS-DA S-plots comparing the 1D ^{1}H NMR spectra obtained from *M. smegmatis* **a)** TAM23 and mc^{2}155, **b)** GPM292 and mc^{2}155 cell lysates. Each point in the S-plot represents a specific bin of integrals for a chemical shift range of 0.025 ppm, where the points at the extreme ends of the S-plot are labeled and are major contributors to the class distinction.

Table 9.5. *M. smegmatis alr* mutants and wild type metabolites in absence of D-alanine

Increase		**Decrease**	
TAM23	**GPM292**	**TAM23**	**GPM292**
Acetate	Alanine	Alpha-ketoisovalerate	Acetate
Alanine	Aspartate	Ethanol	Ethanol
ATP	ATP	Fatty Acid	Fatty Acid
GABA	Glycolate	Glutamate	Glutamate
Glucarate	Glucarate	Glycerol	Glycerol
Glycolate	Lactate	Isoleucine	Isoluecine
Lysine	Lysine	Lactate	NAc-glucosamine-6-phosphate
Phosphoserine	Phosphoserine	NAc-glucosamine-6-phosphate	Oxaloacetate
Serine	Serine	Oxaloacetate	Succinate
Succinate	Succinate	Succinate	Valerate
Threonine	Threonine	Valerate	
UDP	UDP		

The global NMR metabolomics experiment was repeated using *M. smegmatis* mc^2155, TAM23, and GPM292 cells grown in MADC supplemented with 50 mM D-alanine. Again, the PCA 2D scores plot contained three distinct clusters indicating a different metabolome between the *M. smegmatis* wild-type and alanine racemase mutant strains supplemented with D-alanine (Figure 9.9a). The scores plot indicated that PC1 and PC2 corresponded to 26.1% and 14.7% of the variation in the data, respectively. More importantly the three clusters were primarily separated along PC1 suggesting the separation was dominated by the inactivation of alanine racemase, while inherent variations within each group contributed to the variations in PC2.

Similarly, the OPLS-DA 2D scores plot exhibited a similar class separation between *M. smegmatis* mc^2155 wild-type cells and the two alanine racemase mutants (Figure 9.9b). The OPLS-DA yielded a reliable fit as evident by R^2X of 0.559, R^2Y of 0.926, and a Q^2 of 0.862. Importantly, cross validation using CV-Anova yielded a statistically significant p-value of $2.05e^{-10}$, validating the class separation between the two alanine racemase mutants and the mc^2155 wild-type cells. Again, TAM23 and GPM292 formed distinct clusters along the orthogonal component, indicating a different metabolome between the two strains in presence of D-alanine. The S-plot from OPLS-DA was used to identify the bins (chemical shifts) that significantly contributed (0.1 or -0.1) to the class separation (Figure 9.10). The assigned metabolites are summarized (Table 9.6). The pairwise comparison between the alanine racemase insertion (TAM23) and deletion (GPM292) mutants to wild-type mc^2155 cells showed some remarkable

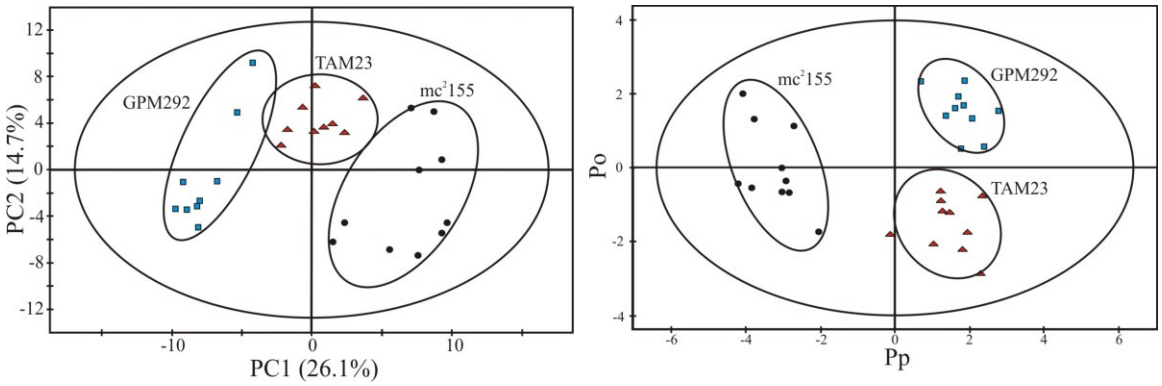


Figure 9.9. a) PCA 2D scores plot and **b)** OPLS-DA 2D scores plot comparing the 1D ^{1}H NMR spectra obtained from lysates of *M. smegmatis* mc^{2}155, TAM23, and GPM292cells grown in MADC media supplemented with 50 mM D-alanine. The wild-type *M. smegmatis* mc^{2}155 cells were designated the control class in the OPLS-DA, and the remainder of the cells was designated as the mutant class. The OPLS-DA used one predictive component and one orthogonal component to yield a R^2X of of 0.559, R^2Y of 0.926, and a Q^2 of 0.862. Cross validation using CV-Anova yielded a p-value of $1.205e^{-10}$. The ellipses correspond to the 95% confidence limits from a normal distribution for each cluster.

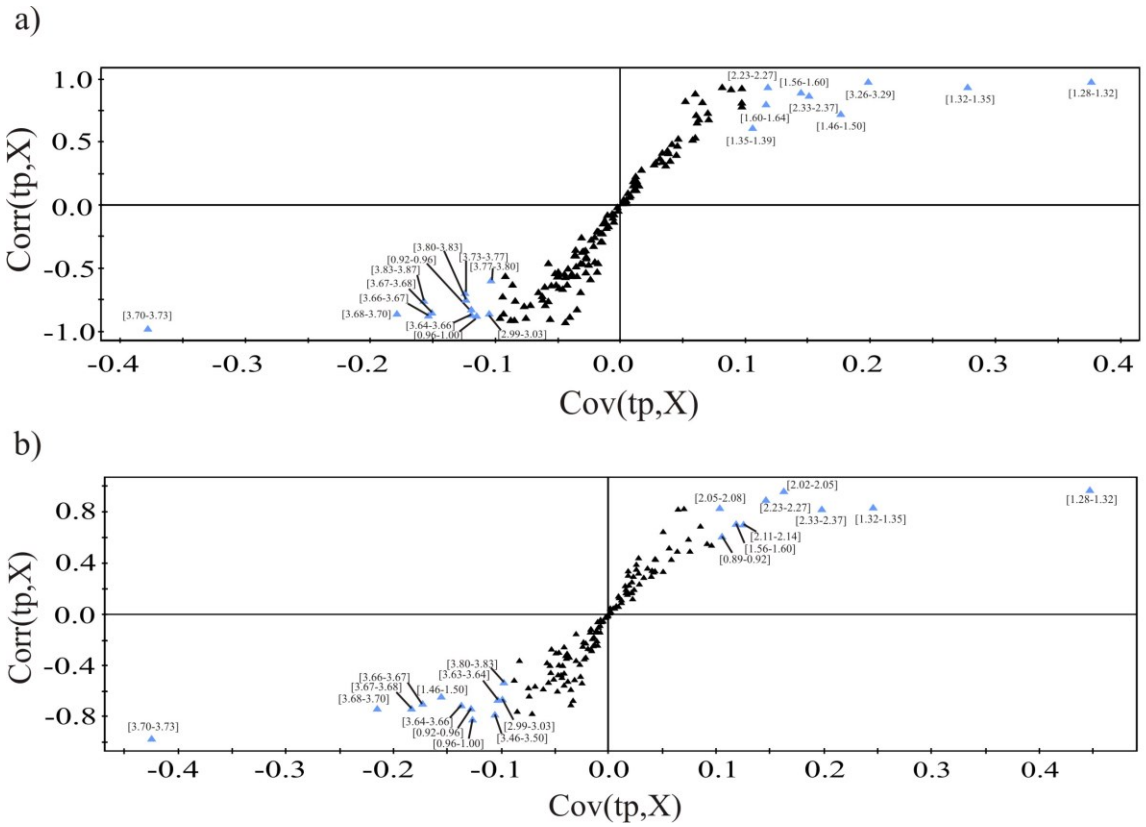


Figure 9.10. OPLS-DA S-plots comparing the 1D ^{1}H NMR spectra obtained from *M. smegmatis* cell lysates **a)** TAM23 and mc^{2}155 **b)** GPM292 and mc^{2}155. All cells were grown in MADC supplemented with 50 mM D-alanine.

Table 9.6. *M. smegmatis alr* mutants and wild type metabolites in presence of 50 mM D-alanine.

Increase		**Decrease**	
TAM23	**GPM292**	**TAM23**	**GPM292**
Citrulline	Alanine	Alanine	Alpha-ketoglutarate
Fatty Acid	Citrulline	Alpha-ketoglutarate	Glucose
Glutamate	Fatty Acid	GABA	GABA
Isoleucine	Glutamate	Lysine	Lysine
Oxaloacetate	Isoleucine	Ornithine	Ornithine
Succinate	Lactate	Trehalose	Trehalose
Threonine	Oxaloacetate	Valine	Valine
	Succinate		
	Threonine		
	Ureidopropionate		

similarities. Metabolites that were increased in TAM23 and GPM292 included glutamate, TCA cycle intermediates, isoleucine, threonine, and compounds with fatty acyl side chains. These metabolites were also identified in the absence of D-alanine (Table 9.5). Metabolites that were decreased in TAM23 and GPM292 included trehalose, lysine and valine. Importantly, the changes in alanine and glucose were different between the alanine racemase mutant strains, where alanine was increased in GPM292, but decreased in TAM23. Conversely, glucose was decreased in GPM292, but did not appear to change significantly in TAM23. Additionally, lactate and ureidopropionate were increased in GPM292, but unaffected in TAM23.

9.3.5 Quantitative analysis of metabolite changes in alanine racemase mutants

The NMR metabolomics experiment was repeated using *M. smegmatis* mc^2155, TAM23, and GPM292 cells grown in MADC supplemented with 100 μM of ^{13}C-labeled D-alanine. ^{13}C-labeled metabolite extracts from three sets of each strain, mc^2155, TAM23, and GPM292 were analyzed using time-zero ^{1}H-^{13}C HSQC ($HSQC_0$) experiments [35, 46]. The 2D ^{1}H-^{13}C HSQC NMR spectra allows for easier identification of metabolites because of the reduced complexity and spectral overlap compared to the 1D ^{1}H NMR spectra. Also, the coupled ^{1}H and ^{13}C chemical shifts obtained for each C-H pair in the 2D ^{1}H-^{13}C HSQC spectra increases the accuracy in metabolite assignments. The 2D ^{1}H-^{13}C HSQC experiment was performed to monitor the flow of carbon-13 throughout the *M. smegmatis* metabolome, where only metabolites that originate from the supplemental ^{13}C-D-alanine are observed (Figure 9.11). Each peak in the 2D ^{1}H-^{13}C

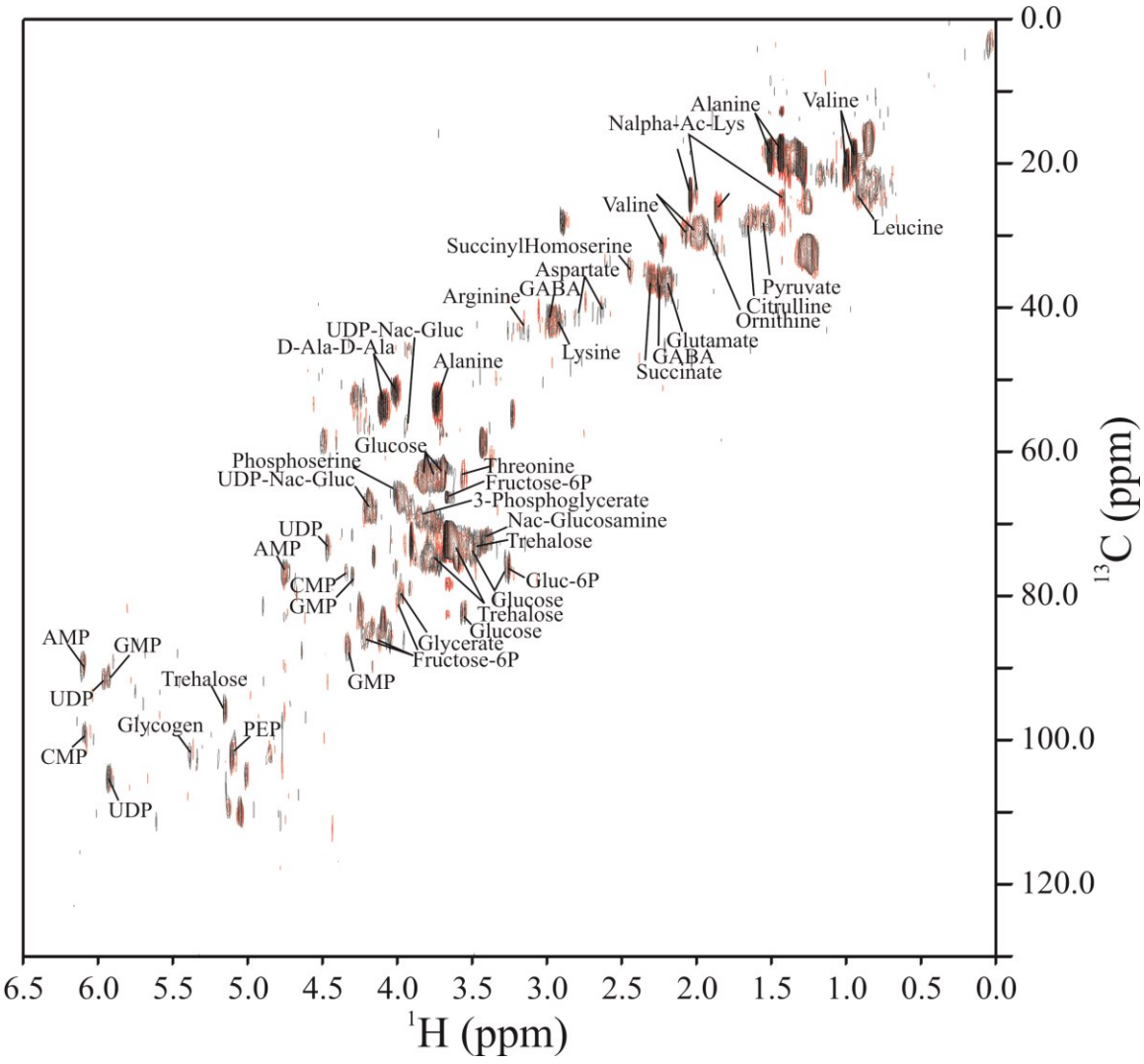


Figure 9.11. Overlay of 2D ^{1}H-^{13}C HSQC spectra. Comparison of cell lysates of *M. smegmatis* TAM23 (black) and GPM292 (red) supplemented with 100 µM ^{13}C-D-alanine.

HSQC spectra represents a ^{13}C-H pair for a specific metabolite that is labeled by matching the chemical shifts against NMR metabolomic databases. A total of 38 metabolites were identified from the three *M. smegmatis* strains. Importantly, the distribution of carbon-13 throughout the *M. smegmatis* metabolome in the two alanine racemase mutants suggested that there is an alternate metabolic pathway involving D-alanine. The metabolite concentrations derived from ^{13}C-D-alanine were compared between the *M. smegmatis* mc^{2}155 wilt-type cells and the two alanine racemase mutants. A concentration for each metabolite was determined by extrapolating peak volumes and intensities from a series of three 2D ^{1}H-^{13}C HSQC spectra at different time points to interpolate a time-zero peak volume or intensity. The spectra were collected in triplicate to determine the average and standard deviation. The average extrapolated peak volume for each metabolite is directly proportional to a concentration based on a calibration curve using 9 ^{13}C-labeled compounds of known concentration. These results permit a direct comparison of metabolite concentration changes between the three *M. smegmatis* strains (Figure 9.12). For example, an 8-fold increase in alanine was observed for GPM292 compared to mc^{2}155, substantially larger than the increase observed in TAM23 compared to the wild type strain. Peptidoglycan precursors such as glutamate, lysine, N-acetyl-glucosamine all decreased significantly in the alanine racemase mutants, as well as the concentration of trehalose, an essential precursor for trehalose-containing glycolipids, which allows the attachment of glycolipids to arabinogalactan in the cell wall [47].

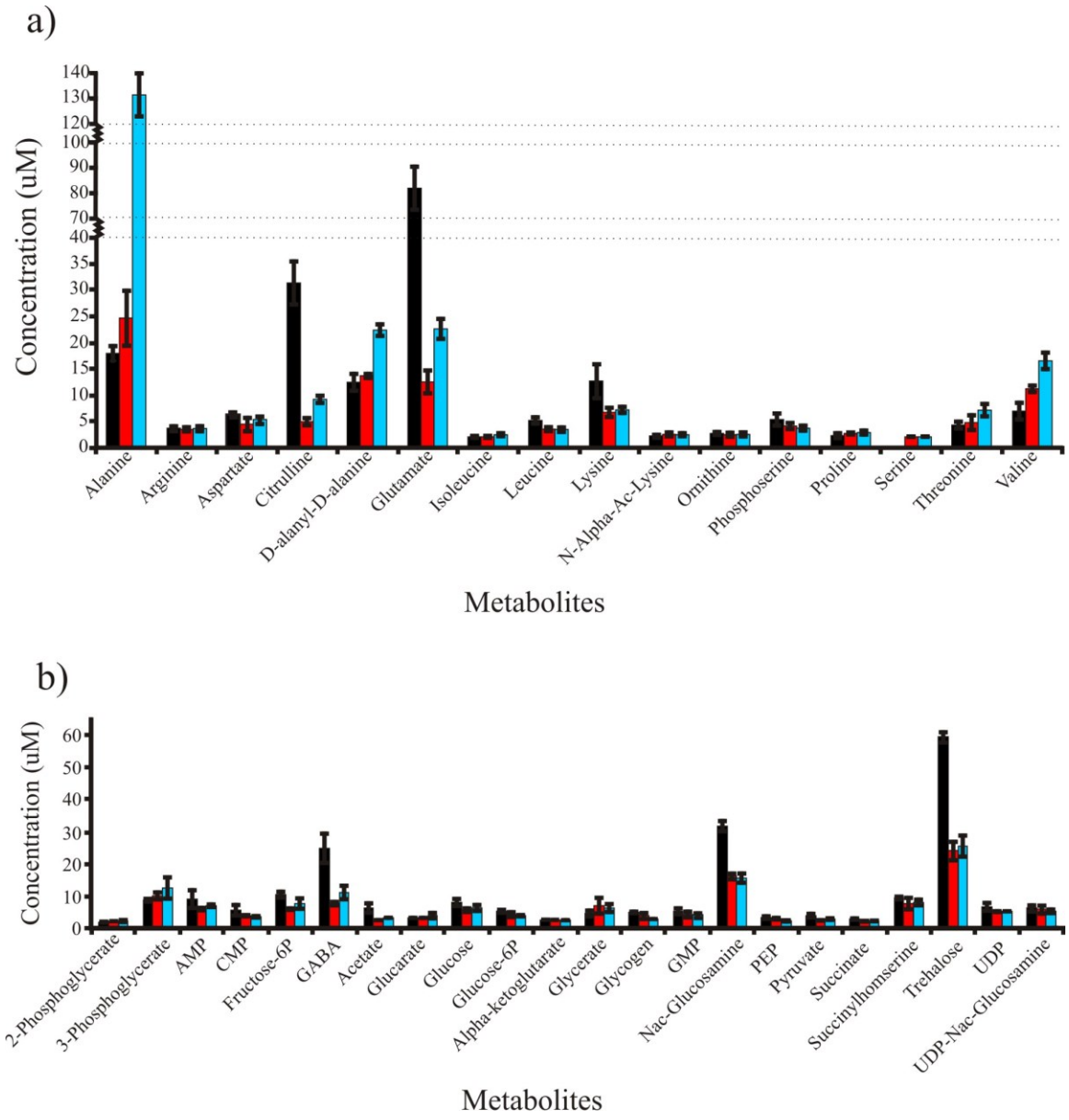


Figure 9.12. Metabolite concentrations. Bar graphs showing the concentrations of metabolites identified in cell lysates from *M. smegmatis* mc^{2}155 (black), TAM23 (red), and GPM292 (blue) using the time-zero 2D ^{1}H-^{13}C HSQC experiment.

9.3.6 Construction, verification and characterization of a M. smegmatis D-alanine auxotroph.

A total of 12,744 mutants were selected after 17 independent transductions of *M. smegmatis* TAM23 with mycobacteriophage phAE159::Tn*53711* (Table 9.2). Initial D-alanine dependence tests, carried out in 3,723 colonies, identified one colony absolutely dependent on D-alanine for growth: *M. smegmatis* TAM23-12. Growth of *M. smegmatis* mc^2155, TAM23 and TAM23-12 in MADC agar with 50 mM D-alanine was detectable after 3 days of incubation. As expected, *M. smegmatis* mc^2155 was able to grow in absence of D-alanine at day 3 of incubation. Similarly, *M. smegmatis* TAM23 was also able to grow in absence of D-alanine although at day 5 of incubation colonies were smaller when compared with the wild type strain (Figure 9.13). *M. smegmatis* TAM23-12 was not able to grow in the absence of D-alanine. Bioinformatic analysis of the transposition insertion identified *murI* as the target gene in TAM23-12 (Figure 9.14). Identity of the gene in TAM23-12 was confirmed by PCR amplification using specific primers outflanking *murI*. As expected, TAM23-12 yielded a 3.9 kb fragment that was 2.2 kb longer than the 1.7 kb wild type mc^2155 product. Complementation using pBUN373 restored *M. smegmatis* TAM23-12 growth in media without D-alanine (data not shown).

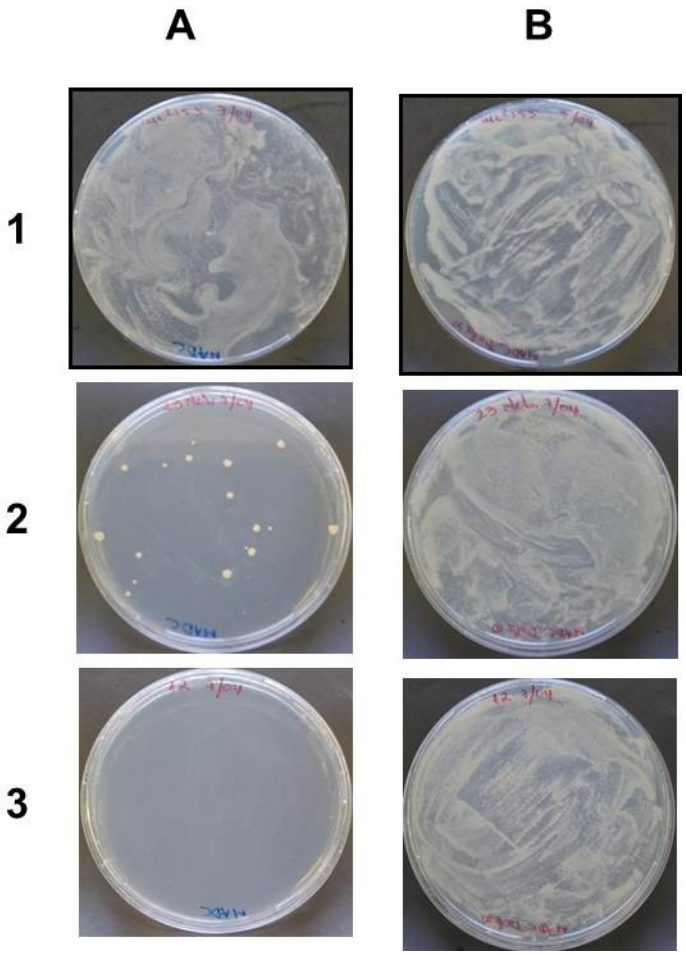


Figure 9.13. D-alanine dependency of *M. smegmatis alr murI* double mutant strain. *M. smegmatis* wild type mc^2155 (1), *alr* mutant TAM23 (2) and double *alr murI* mutant TAM23-12 (3), grown in MADC agar in absence (A) or presence (B) of 50mM D-alanine.

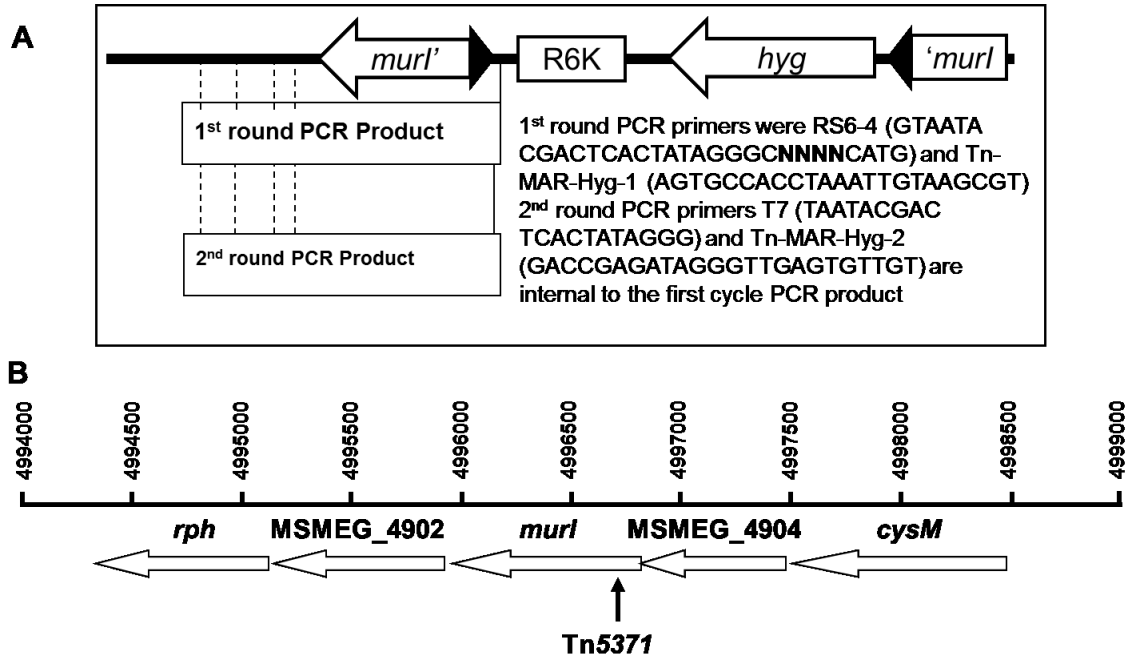


Figure 9.14. Strategy to determine transposon insertion sites. Insertion of Tn*5371* within the *murI* gene is depicted (Panel A) with the corresponding transposon elements: inverted repeats (triangles), R6K origin of replication (open rectangle), and the hygromycin resistant gene (*hyg*). Binding sites for the primers of the first round amplification are shown as vertical dashed (corresponding to various potential locations of the degenerate primer binding sites) or solid lines. Panel B illustrates the location of the insertion site of Tn*5371* identified in TAM23-12.

9.3.7 Construction, verification and characterization of M. tuberculosis strains with inactivation of the chromosomal alr gene.

Our numerous attempts to inactivate the *alr* gene in *M. tuberculosis* H37Rv were unsuccessful. A different strategy was then implemented and extra copies of the *alr* were introduced in *M. tuberculosis* H37Rv before re-attempting the inactivation. The recombinant *M. tuberculosis* H37Rv(pBUN307) strain was confirmed by verifying the presence of pBUN307 plasmid DNA (data not shown), and transfected with phBUN1. Successful inactivation of the chromosomal *alr* gene was confirmed by Southern blot analysis comparing the expected bands (Figure 9.15a) with the observed patterns (Figure 9.15b). The chromosomal *alr* mutant strain with extra copies of *alr* in pBUN307 was named TBcol1. The wild type strain *M. tuberculosis* H37Rv displayed a pattern of 2 bands of 0.675 kb (also present in TBcol1) and 1,682 kb. Besides the 0.675 kb band, the pattern in TBcol1 included two more bands of 0.9 kb (also observed in pBUN288) and 1.522 kb (also found, as a unique band, in pBUN307). The pattern in pBUN288 included another band of 2.634 kb.

Kinyoun staining for acid-alcohol resistant bacilli revealed no differences between wild type, wild type merodiploid or *alr* merodiploid mutant bacilli. After two weeks incubation at 37°C, microscopic examination revealed more dense microcolonies in TBcol1 than in the wild type strains. Macroscopic morphology revealed that wild type colonies were irregular, raised and wrinkled with more dense growth at the center forming a classic cord and less dense and flat borders. In contrast, TBcol1 colonies were irregular, raised, wrinkled and compact (Figure 9.16).

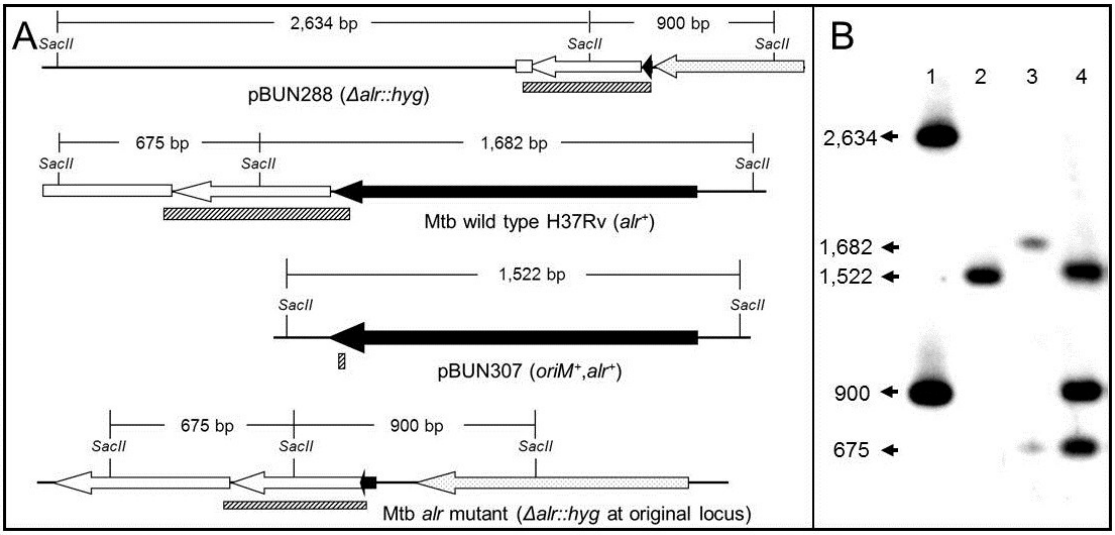


Figure 9.15 . Southern blot analysis of *M. tuberculosis* (Mtb) *alr* mutant for strain H37Rv. **a)** Relevant *Sac*II fragments for plasmid pBUN288, wild type Mtb H37Rv *alr* chromosomal locus, pBUN307, and Mtb *alr* deletion mutant chromosomal locus. Relevant features are *alr* downstream sequences adjacent to the *alr* chromosomal locus (open boxes), *alr* gene (black boxes), 624 bp MTBalrdown probe sequences (striped boxes), and the *hyg* marker (dotted boxes). Different diagrams are not necessarily drawn to scale. **b):** Southern blot. DNA was digested with *Sac*II. MTBalrdown (624 bp) was radiolabelled and used as a probe. **Lane 1**, pBUN288; **Lane 2**, pBUN307; **Lane 3,** Mtb H37Rv; and **Lane 4**, Mtb H37Rv*Δalr::hyg*-pBUN307(*alr*+) (Tbcol1). Fragment sizes (bp) are indicated. Bands of 2,634 and 900 were present in pBUN288 (Lane 1). Plasmid pBUN307 displayed a band of 1,522 (Lane 2). Wild type Mtb H37Rv displays a pattern of two bands of 675 bp and 1,682 bp (Lane 3). Mtb H37Rv merodiploid strain infected with phBUN1 displays a pattern of three bands, the 1,522 bp band from pBUN307 and the 900 bp and 675 bp bands corresponding to the mutated chromosomal *alr* gene. (Lane 4).

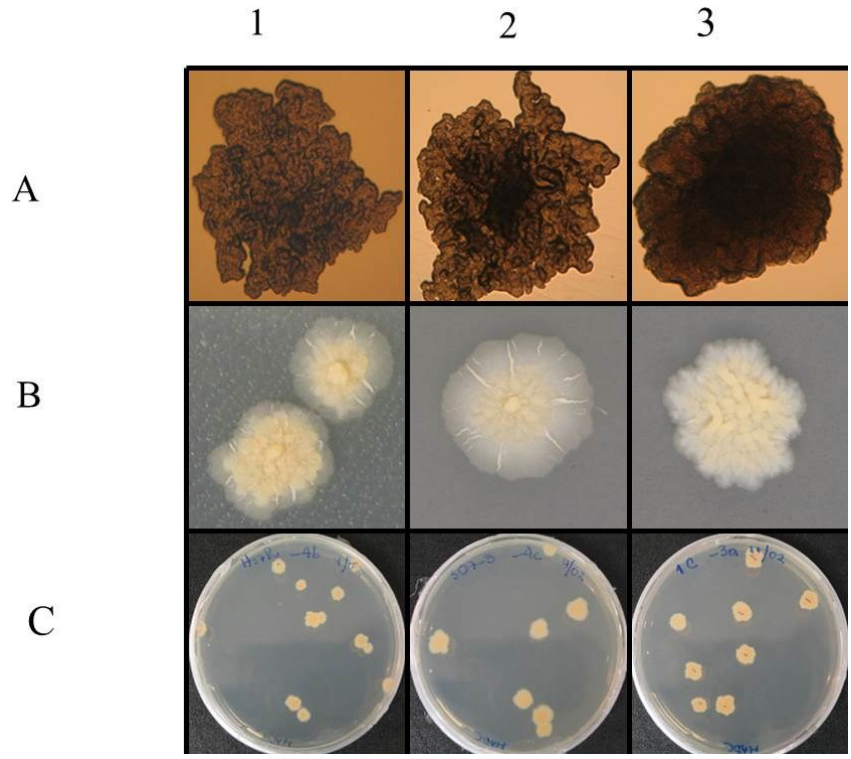


Figure 9.16. Colony morphology of *M. tuberculosis alr* wild type, *alr* merodiploid, and chromosomal *alr* mutant strains. *M. tuberculosis* H37Rv (1), H37Rv(pBUN307) merodiploid (2) and *alr* mutant TBcol1 (3) strains were plated in MADC agar plates supplemented with 50mM D-alanine and incubated at 37ºC.. Individual colonies were compared by microscopy at two weeks (A) and macroscopically at 4 weeks (B and C).

9.4 Discussion

As expected, modifications of the mycobacterial cell wall have been associated with attenuation and increased vulnerability to external agents and conditions [9, 48]. Through the study of mutant strains, this structure has been recognized as an excellent target for antimicrobial agents and a rich source of potential vaccine strains. However, when designing mutant studies related to cell wall biosynthesis, care must be taken to avoid experimental conditions that may lead to inaccurate interpretations. It is known, for example, that *in vitro* conditions not always mimic the in *vivo* environment and that certain media components, such as Malachite Green, may have antimicrobial effects in the presence of an altered cell wall [48].

Given the importance of D-alanine in peptidoglycan biosynthesis [9], the unique *alr* gene in the *M. smegmatis* and *M tuberculosis* genomes was expected to be required in both species for growth in media devoid of D-alanine. However, our previous studies indicated that the *M. smegmatis alr* gene was not required in absence of D-alanine, supporting the existence of an alternative pathway of D-alanine biosynthesis in that species [10, 49]. Our findings were further supported by metabolomics analysis confirming significant differences between the wild type and the *alr* mutant metabolomes [50]. Controversy appeared when *alr* auxotrophy for D-alanine was reported by Milligan et *al* [11] in DM22, a cleverly designed *M. smegmatis alr* deletion mutant. We generated our own deletion mutant, *M. smegmatis* GPM292, in a similar fashion and found that, as our *M. smegmatis* insertional mutant (TAM23), exhibited D-alanine auxotrophy in LBT, but did not require D-alanine in MADC agar. These results indicate that the auxotrophy

conferred by the *alr* mutation depends on the testing media. The growth patterns observed in different solid media identified casein hydrolysate as the responsible component for growth inhibition in LBT and Middlebrook 7H11. Previous reports have associated susceptibility to Malachite Green, a synthetic dye and component of 7H10 and 7H11, to cell wall defects [48]. However, Malachite Green did not inhibit growth of *M. smegmatis alr* mutants as both *M. smegmatis* TAM23 and GPM292 grew well in 7H10, a medium that contains the dye but lacks casein hydrolysate. Therefore, we concluded that the lack of growth observed in *M. smegmatis alr* mutants in 7H11 or LBT was likely due to the presence of free amino acids provided by the casein hydrolysate. Further testing attributed the inhibitory effect to the presence of L-alanine, L-serine, L-asparagine, or L-glutamine. Additional studies are warranted to identify the role of these metabolites in growth inhibition.

D-alanine can be synthesized in bacteria through a transamination reaction of the general type: D-amino acid + pyruvate → α-keto-acid + D-alanine, catalyzed by D-amino acid aminotransferase, encoded by the *dat* gene. In the case of the intracellular pathogen *Listeria monocytogenes*, inactivation of both *alr* and *dat* genes is required to achieve D-alanine auxotrophy [51]. Unfortunately, a *dat* homologue has not been identified in mycobacteria. However, the existence of an unknown transaminase in *M. smegmatis* is supported by the growth of both *M. smegmatis alr* mutants to levels similar to the wild type strain in media with D-glutamic acid and pyruvate, but without D-alanine. Likewise, identification of the glutamate recemase (*murI*) mutation as responsible for complete D-alanine auxotrophy in TAM23-12 further supports the involvement of D-glutamate in the

alternative pathway of D-alanine biosynthesis in *M. smegmatis.* Our NMR studies revealing decreased glutamate levels in absence of D-alanine and increased levels in its presence, compared to the wild type glutamate levels, further strengthens the case for this alternative D-alanine biosynthetic route. In summary, the *M. smegmatis alr* gene is non-essential and its inactivation confers conditional D-alanine auxotrophy. Otherwise, the *alr* gene is dispensable in absence of D-alanine. Thus, a regulated pathway of D-alanine biosynthesis, dependent on MurI and a novel but still unknown D-amino acid aminotransferase, underlies this phenotype.

For *M. tuberculosis* H37Rv, our study indicates that Alr is an essential function in H37Rv, as gene inactivation requires extra wild type copies of the *alr* gene in either the presence or absence of D-alanine in MADC media. These results differ from those of Awasthy *et al*. [12] that were able to obtain H37Rv *alr* mutants without providing extra copies of the gene, by selecting mutants on MADC media supplemented with D-alanine. Their results indicate that Alr is not essential in *M. tuberculosis* since its inactivation leads to auxotrophy that can be compensated by a low concentration of D-alanine (e. g., 10 µg/ml, approx 0.1 mM). There are several possibilities that could explain these seemingly contradictory results. The mutants isolated by Awasthy *et al* could have acquired, during the lengthy selection process in the sucrose selection media, a compensatory mutation that partially activates the putative alternative pathway of D-alanine biosynthesis that we discovered in *M. smegmatis*. In that case, this activation would not be sufficient to allow for growth in complete absence of D-alanine, but allow the *alr* mutants to grow with very little D-alanine. Otherwise, this low requirement would

be counterintuitive as mycobacteria incorporate D-alanine in both peptidoglycan and glycolipids [49]. In that case, it could be further argued that additional compensatory mutations may occur both *in vitro* and *in vivo* that could lead to activation of the alternative pathway at sufficient levels so as to allow growth in complete absence of D-alanine. Compensatory mutations leading to increased virulence *in vivo* have been reported for isoniazid resistant strains. Moreover, Aswasthy *et al*., noted that *alr* mutants were able to persist at low levels in mice.

Alternatively, the corresponding H37Rv laboratory strains used in the two studies may differ in genetic regions relevant to D-alanine biosynthesis and uptake. In this context, these isolates correspond to different American Type Culture Collection Catalogue numbers (ATCC 25,698 in our study versus ATCC 27,294 in Awasthy *et al*.). It is to be noted that genomic differences have been reported in H37Rv laboratory strains even among isolates derived from the same ATCC substrain. Lastly, it is also possible that the stringent selection used in our shuttle plasmid gene inactivation procedure, to select directly for double cross over events, inhibits the growth of any potential *alr* mutant strains at a critical time. Discovery of all genes involved in the alternative pathway of D-alanine biosynthesis in *M. smegmatis* and the presence or absence of homologues in *M. tuberculosis* may help to sort out these alternative hypotheses.

Regarding the lethal target(s) of DCS in mycobacteria, our studies suggest that Alr is not the lethal target in *M. smegmatis*, as *alr* mutants can grow in absence of D-alanine. The inhibitory effect of DCS on *M. smegmatis* is likely due to another target such as D-alanine ligase [52]. Alternatively, DCS may be able to also inhibit the

alternative pathway of D-alanine biosynthesis and thus determine a complete shut-down of this biosynthetic process. A similar situation may or may not occur in *M. tuberculosis*, depending upon how these possibilities described above are sorted out. In that case, if the genes required for the alternative pathway are partially or fully absent from *M. tuberculosis*, then Alr is likely to be the primary target of DCS.

Our long term goal is to apply our studies to design novel therapeutic agents and/or attenuated vaccine strains that can be used to control human TB. Thus, our findings should be interpreted in the context of the in *vivo* interaction between the pathogen and its human host environment. The ability to survive inside human macrophages constitutes a major factor in *M. tuberculosis* pathogenesis. Therefore, understanding the dynamic interaction between *M. tuberculosis* and the host cell is of paramount importance in the development of new, more effective tools to prevent and control TB. Although this issue has been addressed using cleverly designed *in vitro* and *in vivo* models of infection, a major drawback is the lack of an appropriate matrix that permits their interpretation in the context of pathogen-cell interactomes. Fortunately, completion of the *M. tuberculosis* and human genomes and recent bioinformatics advances have enabled the development of promising genome-wide metabolic reconstructions [30, 53-56] that represent important tools to achieve a comprehensive understanding of the dynamic interactions between *M. tuberculosis* and the host cell(s).

In the context of our study, D-alanine has been identified in human tissues and a potential role in neurotransmission has been suggested [57-60]. Free D-alanine (9.52 nmol/g) has been detected in the brain of normal subjects while higher levels (20.8

nmol/g) were detected in Alzheimer's patients [60]. Moreover, D-alanine is the substrate for a human D-amino acid oxidase, an enzyme that has been associated with increased susceptibility to schizophrenia. The affinity of the enzyme for D-alanine is higher than for its putative substrate, D-serine (Km, approx.1.3 mM and 7.5 mM, respectively) [61]. Furthermore, oral administration of D-alanine (100 mg/kg/day) to schizophrenic patients resulted in significant clinical improvement without major side effects, making this amino acid a promising candidate for schizophrenia pharmacotherapy [62]. The high affinity of a human enzyme for D-alanine, coupled with the side effects of the D-alanine analog D-cycloserine on the nervous system, indicate that D-alanine may play an important role in normal subjects and in patients with certain neurological conditions [63]. Therefore, further studies on the specific role of D-alanine in humans are warranted in order to maximize the efficacy and safety of novel antimicobacterial tools based on *alr* inactivation. The use of those tools might be contraindicated in the presence of certain medical conditions or pharmacotherapy agents. Thus, the success of any application that is dependent on the inhibition of D-alanine biosynthesis in *M. tuberculosis* must be evaluated with the assumption that low concentrations of D-alanine are found in human tissues. In that context, the results of Awasthy *et al.* indicate that *M. tuberculosis alr* mutants may be inappropriate as TB vaccine candidates as they could grow to a considerable extent in human tissue regardless of the presence or absence of an alternative pathway of D-alanine biosynthesis in *M. tuberculosis*. Likewise, Alr inhibitors would not be effective to treat TB as the low concentrations of D-alanine could compensate for *alr* gene inactivation. In summary, this study indicates that several

factors ranging from experimental design to result interpretation, and comprehensive knowledge and understanding of the specific interaction between the *Mycobacterium* and the host cell, are extremely important for a successful use of *alr* mutants in the design of novel tools urgently needed to control TB.

9.5 References

1. World Health Organization, *Global Tuberculosis Report 2012*, 2012, World Health Organization: Geneva, Switzerland:. p. p. 100.

2. Brennan, P.J. and G.S. Besra, *Structure, Function and Biogenesis of the Mycobacterial Cell Wall.* Biochemical Society Transactions, 1997. **25**(1): p. 188-194.

3. Moat, A., Foster, J., Spector, M., ed. *Cell Structure and Function*. ed. M. Physiology2002, Wiley-Liss and John Wiley & Sons: New York. 277-349.

4. Belanger, A.E., J.C. Porter, and G.F. Hatfull, *Genetic Analysis of Peptidoglycan Biosynthesis in Mycobacteria: Characterization of a ddlA Mutant of Mycobacterium smegmatis.* Journal of Bacteriology, 2000. **182**(23): p. 6854-6856.

5. Neuhaus, F.C., *Selective Inhibition of Enzymes Utilizing Alanine in the Biosynthesis of Peptidoglycan.* Antimicrobial Agents and Chemotherapy, 1967. **7**: p. 304-13.

6. Lambert, M.P. and F.C. Neuhaus, *Factors Affecting the Level of Alanine Racemase in Escherichia coli.* Journal of Bacteriology, 1972. **109**(3): p. 1156-61.

7. Walsh, C.T., *Enzymes in the D-Alanine Branch of Bacterial Cell Wall Peptidoglycan Assembly.* The Journal of Biological Chemistry, 1989. **264**(5): p. 2393-6.

8. van Heijenoort, J., ed. *Biosynthesis of Bacterial Peptidoglycan*. Bacterial Cell Wall, ed. J.H.R. Ghuysen1994, Elsevier Medical Press: Amsterdam. 39-54.

9. Brennan, P.J., *Structure, Function, and Biogenesis of the Cell Wall of Mycobacterium tuberculosis.* Tuberculosis, 2003. **83**(1-3): p. 91-7.

10. Chacon, O., et al., *Mycobacterium smegmatis D-Alanine Racemase Mutants are Not Dependent on D-Alanine for Growth.* Antimicrobial Agents and Chemotherapy, 2002. **46**(1): p. 47-54.

11. Milligan, D.L., et al., *The Alanine Racemase of Mycobacterium smegmatis is Essential for Growth in the Absence of D-Alanine.* Journal of Bacteriology, 2007. **189**(22): p. 8381-6.

12. Awasthy, D., et al., *Alanine Racemase Mutants of Mycobacterium tuberculosis Require D-Alanine for Growth and are Defective for Survival in Macrophages and Mice.* Microbiology, 2012. **158**(Pt 2): p. 319-27.

13. Krzywinska, E., J. Krzywinski, and J.S. Schorey, *Naturally Occurring Horizontal Gene Transfer and Homologous Recombination in Mycobacterium.* Microbiology, 2004. **150**(Pt 6): p. 1707-12.

14. Veyrier, F., et al., *Phylogenetic Detection of Horizontal Gene Transfer During the Step-Wise Genesis of Mycobacterium tuberculosis.* BMC Evolutionary Biology, 2009. **9**: p. 196.

15. Snapper, S.B., et al., *Isolation and Characterization of Efficient Plasmid Transformation Mutants of Mycobacterium smegmatis.* Molecular Microbiology, 1990. **4**(11): p. 1911-9.

16. Caceres, N.E., et al., *Overexpression of the D-alanine Racemase Gene Confers Resistance to D-cycloserine in Mycobacterium smegmatis.* Journal of Bacteriology, 1997. **179**(16): p. 5046-5055.

17. Connell, N.D., et al., *Effective Immunization Against Cutaneous Leishmaniasis with Recombinant Bacille Calmette-Guerin Expressing the Leishmania Surface Proteinase gp63.* Proceedings of the National Academy of Sciences of the United States of America, 1993. **90**(24): p. 11473-7.

18. Pavelka, M.S., Jr. and W.R. Jacobs, Jr., *Comparison of the Construction of Unmarked Deletion Mutations in Mycobacterium smegmatis, Mycobacterium Bovis bacillus Calmette-Guerin, and Mycobacterium tuberculosis H37Rv by Allelic Exchange.* Journal of Bacteriology, 1999. **181**(16): p. 4780-9.

19. Kriakov, J., S. Lee, and W.R. Jacobs, Jr., *Identification of a Regulated Alkaline Phosphatase, A Cell Surface-Associated Lipoprotein, in Mycobacterium smegmatis.* Journal of Bacteriology, 2003. **185**(16): p. 4983-91.

20. Bardarov, S., et al., *Conditionally Replicating Mycobacteriophages: A System for Transposon Delivery to Mycobacterium tuberculosis.* Proceedings of the National Academy of Sciences of the United States of America, 1997. **94**(20): p. 10961-6.

21. Bardarov, S., et al., *Specialized Transduction: An Efficient Method for Generating Marked and Unmarked Targeted Gene Disruptions in Mycobacterium tuberculosis, M. bovis BCG and M. smegmatis.* Microbiology, 2002. **148**(Pt 10): p. 3007-17.

22. Braunstein, M., S.S. Bardarov, and W.R. Jacobs, *Genetic Methods for Deciphering Virulence Determinants of Mycobacterium tuberculosis.* Bacterial Pathogenesis, Pt C, 2002. **358**: p. 67-99.

23. Sambrook, J.R., D., *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press: p. 1.38-9.1.

24. Zwietering, M.H., et al., *Modeling of the Bacterial Growth Curve.* Applied and Environmental Microbiology, 1990. **56**(6): p. 1875-81.

25. Hwang, T., Shaka, A., *Water Suppression that Works. Excitation Sculpting using Arbitrary Waveforms and Pulsed Field Gradients.* Journal of Magnetic Resonance, 1995. **112**: p. 275-279.

26. Hu, K., W.M. Westler, and J.L. Markley, *Simultaneous Quantification and Identification of Individual Chemicals in Metabolite Mixtures by Two-Dimensional Extrapolated Time-Zero (1)H-(13)C HSQC (HSQC(0)).* Journal of the American Chemical Society, 2011. **133**(6): p. 1662-5.

27. Zhang, B., et al., *NMR Analysis of a Stress Response Metabolic Signaling Network.* Journal of Proteome Research, 2011. **10**(8): p. 3743-54.

28. Delaglio, F., et al., *NMRPipe: A Multidimensional Spectral Processing System Based on UNIX Pipes.* Journal of Biomolecular NMR, 1995. **6**(3): p. 277-93.

29. Johnson, B.A., *Using NMRView to Visualize and Analyze the NMR Spectra of Macromolecules.* Methods in Molecular Biology, 2004. **278**: p. 313-52.

30. Wishart, D.S., et al., *HMDB: A Knowledgebase for the Human Metabolome.* Nucleic Acids Research, 2009. **37**(Database issue): p. D603-10.

31. Cui, Q., et al., *Metabolite Identification via the Madison Metabolomics Consortium Database.* Nature Biotechnology, 2008. **26**(2): p. 162-4.

32. Akiyama, K., et al., *PRIMe: A Web Site that Assembles Tools for Metabolomics and Transcriptomics.* In Silico Biology, 2008. **8**(3-4): p. 339-45.

33. Okuda, S., et al., *KEGG Atlas Mapping for Global Analysis of Metabolic Pathways.* Nucleic Acids Research, 2008. **36**(Web Server issue): p. W423-6.

34. Caspi, R., et al., *The MetaCyc Database of Metabolic Pathways and Enzymes and the BioCyc Collection of Pathway/Genome Databases.* Nucleic Acids Research, 2010. **38**: p. D473-D479.

35. Hu, K., et al., *Measurement of Absolute Concentrations of Individual Compounds in Metabolite Mixtures by Gradient-Selective Time-Zero 1H-13C HSQC with Two Concentration References and Fast Maximum Likelihood Reconstruction Analysis.* Analytical Chemistry, 2011. **83**(24): p. 9352-60.

36. Larsen, M.H., et al., *Genetic Manipulation of Mycobacterium tuberculosis.* Current Protocols in Microbiology, 2007. **Chapter 10**: p. Unit 10A 2.

37. McAdam, R.A., et al., *Characterization of a Mycobacterium tuberculosis H37Rv Transposon Library Reveals Insertions in 351 ORFs and Mutants with Altered Virulence.* Microbiology, 2002. **148**(Pt 10): p. 2975-86.

38. Shin, S.J., et al., *Identification of Novel Virulence Determinants in Mycobacterium paratuberculosis by Screening a Library of Insertional Mutants.* Infection and Immunity, 2006. **74**(7): p. 3825-33.

39. Bolivar, F., et al., *Origin of Replication of Pbr345 Plasmid DNA.* Proceedings of the National Academy of Sciences of the United States of America, 1977. **74**(12): p. 5265-5269.

40. Kalpana, *Towards the Development of Moleculare Genetic Approaches for the Study of Mycobacterial Pathogenesis*, in *Albert Einstein College of Medicine*1991.

41. van Soolingen, D., P.E. de Haas, and K. Kremer, *Restriction Fragment Length Polymorphism Typing of Mycobacteria.* Methods in Molecular Medicine, 2001. **54**: p. 165-203.

42. Kent, P., Kubica, G., ed. *Identification Isolation*. Public Health Mycobacteriology: a Guild for the Level III Laboratory, ed. P.H.S.C.f.D. Control1985, Division of Laboratroy Training and Consultation, Laboratroy Program Office: Atlanta, Georgia.

43. Strych, U., et al., *Characterization of the Alanine Racemases From Two Mycobacteria.* FEMS Microbiology Letters, 2001. **196**(2): p. 93-8.

44. Bylesjo, M., et al., *OPLS Discriminant Analysis: Combining the Strengths of PLS-DA and SIMCA Classification.* Journal of Chemometrics, 2006. **20**(8-10): p. 341-351.

45. Eriksson, L., Johansson, E., Kettaneh-Wold, N., Wikstrom, C., and Wold, S., ed. *Design of Experiments-Principles and Applications*. 2008, Umetrics.

46. Hu, K., W.M. Westler, and J.L. Markley, *Simultaneous Quantification and Identification of Individual Chemicals in Metabolite Mixtures by Two-Dimensional Extrapolated Time-Zero (1)H-(13)C HSQC (HSQC(0)).* Journal of the American Chemical Society, 2011.

47. Woodruff, P.J., et al., *Trehalose is Required for Growth of Mycobacterium smegmatis.* Journal of Biological Chemistry, 2004. **279**(28): p. 28835-28843.

48. Banaei, N., et al., *Lipoprotein Processing Is Essential for Resistance of Mycobacterium tuberculosis to Malachite Green.* Antimicrobial Agents and Chemotherapy, 2009. **53**(9): p. 3799-3802.

49. Chacon, O., et al., *Impairment of D-Alanine Biosynthesis in Mycobacterium smegmatis Determines Decreased Intracellular Survival in Human Macrophages.* Microbiology, 2009. **155**(Pt 5): p. 1440-50.

50. Halouska, S., et al., *Use of NMR Metabolomics to Analyze the Targets of D-Cycloserine in Mycobacteria: Role of D-Alanine Racemase.* Journal of Proteome Research, 2007. **6**(12): p. 4608-14.

51. Thompson, R.J., et al., *Pathogenicity and Immunogenicity of a Listeria monocytogenes Strain That Requires D-Alanine for Growth.* Infection and Immunity, 1998. **66**(8): p. 3552-61.

52. Feng, Z. and R.G. Barletta, *Roles of Mycobacterium smegmatis D-Alanine:D-Alanine Ligase and D-Alanine Racemase in the Mechanisms of Action of and Resistance to the Peptidoglycan Inhibitor D-Cycloserine.* Antimicrobial Agents and Chemotherapy, 2003. **47**(1): p. 283-91.

53. Duarte, N.C., et al., *Global Reconstruction of the Human Metabolic Network Based on Genomic and Bibliomic Data.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(6): p. 1777-82.

54. Bordbar, A., et al., *Insight into Human Alveolar Macrophage and M. tuberculosis Interactions via Metabolic Reconstructions.* Molecular Systems Biology, 2010. **6**.

55. Bordbar, A., et al., *A Multi-Tissue Type Genome-Scale Metabolic Network for Analysis of Whole-Body Systems Physiology.* Bmc Systems Biology, 2011. **5**.

56. Bordbar, A. and B.O. Palsson, *Using the Reconstructed Genome-Scale Human Metabolic Network to Study Physiology and Pathology.* Journal of Internal Medicine, 2012. **271**(2): p. 131-141.

57. Bruckner, H. and A. Schieber, *Determination of Amino Acid Enantiomers in Human Rrine and Blood Serum by Gas Chromatography-Mass Spectrometry.* Biomedical Chromatography, 2001. **15**(3): p. 166-172.

58. Zhao, S., Wang, B., He, M., Bai, W., Chen, L., *Determination of Free D-Alanine in the Human Plasma by Capillary Electrophoresis with Optical Fiber Light-Emitting Diode-Induced Fluorescence Detection.* Analyca Chimica Acta, 2006. **569**: p. 182-187.

59. Nagata, K., et al., *L-Serine, D- and L-Proline and Alanine as Respiratory Substrates of Helicobacter pylori: Correlation Between in vitro and in vivo Amino Acid Levels.* Microbiology, 2003. **149**(Pt 8): p. 2023-30.

60. Fisher, G.H., et al., *Free D-Aspartate and D-Alanine in Normal and Alzheimer brain.* Brain Research Bulletin, 1991. **26**(6): p. 983-5.

61. Molla, G., et al., *Characterization of Human D-Amino Acid Oxidase.* FEBS letters, 2006. **580**(9): p. 2358-64.

62. Tsai, G.E., et al., *D-Alanine Added to Antipsychotics for the Treatment of Schizophrenia.* Biological Psychiatry, 2006. **59**(3): p. 230-4.

63. van Berckel, B.N., et al., *D-Cycloserine Increases Positive Symptoms in Chronic Schizophrenic Patients When Administered in Addition to Antipsychotics: A Double-Blind, Parallel, Placebo-Controlled Study.* Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology, 1999. **21**(2): p. 203-10.

CHAPTER 10

METABOLOMICS ANALYSIS IDENTIFIES D-ALANINE-D-ALANINE LIGASE AS THE PRIMARY LETHAL TARGET OF D-CYCLOSERINE IN MYCOBACTERIA

10.1 Introduction

Tuberculosis (TB) remains one of the leading causes of morbidity and mortality from a single infectious disease on a global perspective [1]. In 2011, 8.7 million people were infected with TB with over 1.8 million deaths world-wide. Moreover, the emergence of multiple- (MDR-TB) and extensively drug-resistant- (XDR-TB) *Mycobacterium tuberculosis* strains threaten to curtail efforts in disease control. Currently, about 3.7% of new TB patients are infected with MDR-TB. To achieve the WHO targeted goal of a successful treatment rate of 75% by 2015, a better understanding of the molecular mechanisms of action and resistance to existing antibiotics and the development of novel drugs that are more potent and safer is urgently needed.

D-Cycloserine (DCS) is a second line drug that is currently used as a last resort on MDR- and XDR-TB. DCS has been used to treat TB for over fifty years despite a lack of knowledge regarding the identity of its lethal target [2]. Although DCS inhibits bacterial cell growth, it has serious neurological side effects [3-5]. DCS treatment results in psychosis, depression, and convulsions, among other issues. In this context, DCS has also been tested in neurological studies and has been shown to act as a partial agonist of *N*-methyl-D-aspartate (NMDA) and glycine receptors [6]. Thus, understanding the source

of DCS antimicrobial activity would enable the development of next-generation antibiotics for TB that avoids these CNS side effects. DCS is a cyclic analogue of D-alanine and has been shown to competitively inhibit Alanine racemase (Alr, EC 5.1.1.1) and D-alanine-D-alanine ligase (Ddl, 6.3.2.4) [7, 8]. Correspondingly, the current understanding of DCS activity against mycobacteria is through the inhibition of Alr and/or Ddl that halts the synthesis of peptidoglycan precursors. However, we previously demonstrated that Alr is not the lethal target of DCS, suggesting Ddl is the primary target in live mycobacteria [9]. This is also consistent with the observation by Takayama *et al.* that the UDP-MurNAc-tripeptide accumulates in *M. tuberculosis* upon treatment with DCS.[10] UDP-MurNAc-tripeptide is the product of the MurE meso-diaminopimelate-adding enzyme (EC 6.3.2.13) in the peptidoglycan biosynthesis pathway, while the next step involves the MurF D-alanyl-D-alanine adding enzyme (EC 6.3.2.10) ligating the UDP-MurNAc-tripeptide with D-alanyl-D-alanine, the product of Ddl [11]. Clearly, the inhibition of Ddl by DCS would decrease the production of D-alanyl-D-alanine and lead to the observed accumulation of the UDP-MurNAc-tripeptide, as the co-substrate of the MurF reaction is decreased.

Peptidoglycan biosynthesis is an ideal target for drug design because the pathway is not present in mammalian cells [11]. Also, the resulting peptidoglycan-arabinogalactan complex gives the cell its structural integrity [11-13]. The peptidoglycan layer consists of an alternating N-acetylglucosamine and N-glycolated or N-acetylated muramic acid. Each N-glycolated or N-acetylated muramic acid is bound to a tetrapeptide consisting of an L-alanyl–D-isoglutaminyl–meso-diaminopimelyl–D-alanine (L-Ala–D-Glu–A_2pm–D-

Ala) moiety. The tetrapeptide forms a crosslink between adjacent alternating aminosugars where the side chain of meso-diaminopimelate of one group forms a peptide bond with a D-alanine or meso-diaminopimelate residue from the adjacent group. This large mycolyl-arabinogalactan-peptidoglycan complex creates an impermeable barrier that is essential for the viability of the cell [14]. This metabolic route includes the D-alanine branch pathway consisting of three enzymes that contribute to the synthesis of the glycomuramyl pentapeptide involved in the crosslinking of mature peptidoglycan. Alr is a pyridoxal phosphate dependent enzyme that interconverts L-alanine and D-alanine [15, 16]. The ATP Ddl catalyzes the subsequent peptide bond between two D-alanine moieties [17]. The final step is the addition of this dipeptide to the glycomuramyl tripeptide cytoplasmic precursor by the ATP-dependent MurF adding enzyme to form the pentapeptide complex [18]. As a result, numerous enzymes within the peptidoglycan biosynthesis machinery recognize or bind a D-alanine moiety and are potentially inhibited by DCS. Herein, we describe our application of Nuclear Magnetic Resonance (NMR) based metabolomics and bioinformatics to determine the lethal target of DCS in mycobacteria and to investigate the effects of this drug on central metabolic pathways related to peptidoglycan biosynthesis. Our analysis indicates that DCS is a promiscuous inhibitor targeting multiple enzymes within the peptidoglycan biosynthesis pathway, but Ddl is the primary target. Our results also support the prior observation by Caceres *et al.* [19] that D-alanine is a competitive inhibitor of DCS and its over-production is a primary mechanism of resistance.

10.2 Method and Materials

10.2.1 Preparation of M. tuberculosis NMR metabolomics samples

General procedures for the handling and preparation of *M. tuberculosis* NMR samples for metabolomic analysis have been described elsewhere [20]. Six (3 for ^{13}C-D-alanine and 3 for ^{13}C-D-alanine DCS) *M. tuberculosis* H37Rv 110 mL MADC broth (Middlebrook 7H9 complete media) cultures from glycerol stock were grown shaking at 100 rpm at 37ºC for approximately 7 days (OD_{600} 0.6-0.8). A 500 µL sample was removed and kept at 4ºC for CFU determinations. ^{13}C-D-alanine (0.1 mM final concentration; e.g.,100 µL of a 100 mM stock) was added to all cultures. DCS (50 µg/mL final concentration (500 µL of a 10 mg/mL stock)) was added to only 3 flasks. All cultures were incubated for an additional 18 hours, taking another OD_{600} reading and removing a 500 µL sample from each flask. Cultures were placed on ice for 5 minutes and left on ice throughout the rest of the processing. Cultures were harvested by spinning them down at 2000 g at 4ºC for 15 minutes in 50 mL tubes. The samples were washed two times (~25-30 mL) with ice-cold double distilled water (ddH_2O). The cell pellets were re-suspended with 1 mL of ddH_2O and transferred to a 2 mL vial consisting of 0.1 mm silica beads (Lysing Matrix B). The cells were then lysed using a FastPrep-24 instrument for 60 seconds at 6 m/s. The cellular mixture was centrifuged at 15,000 g at 4ºC for 10 minutes and the supernatant was extracted to a 1.5 mL tube. 700 µL of ddH_2O was added to the tube containing the lysing matrix B and briefly vortexed, followed by centrifugation, and then transferring and combining the supernatants in a 1.5 mL tube. The supernatant was syringe filtered (0.2 µm) into a sterile tube. A 100 µl of the sample

(10%) was plated on MADC to verify that there are no live cells. Samples were frozen in an EtOH-dry ice bath and stored at -80ºC for 2 months. After verification that the plates contained no viable cells (CFUs), the metabolomic samples were taken out of BSL3 containment, lyophilized and prepared for analysis by NMR. 10 µL of the supernatant was used to determine protein concentration as described below for the *M. smegmatis* samples.

10.2.2 Preparation of M. smegmatis NMR metabolomics samples

NMR samples for 2D ^{1}H-^{13}C HSQC experiments were prepared from 6 groups of triplicate and independent *M. smegmatis* $mc^{2}155$ cultures using 100 µM $^{13}C_2$-D-alanine ($^{13}C\alpha$ and $^{13}C\beta$ labeled) or $^{13}C_3$-pyruvate as a carbon-13 source. The following groups using $^{13}C_2$-D-alanine are $mc^{2}155$ as a control and $mc^{2}155$ treated with 75, 300, or 1200 µg/mL of DCS. The groups using $^{13}C_3$-pyruvate are untreated $mc^{2}155$ and $mc^{2}155$ treated with 75 µg/mL DCS. The replicate cultures were grown at 37°C with shaking at 200 rpm in 110 mL of MADC (250 mL flask) until an OD_{600} of 0.6 was met. $^{13}C_2$-D-alanine or $^{13}C_3$-pyruvate was inoculated to the designated cultures for a final concentration of 100 µM. The cultures were allowed to grow for 10 minutes and then were treated with DCS to a final concentration of 75, 300, or 1200 µg/mL. The cultures were then grown for an additional 2 hours before harvesting. Each culture was placed on ice for 5 minutes and then centrifuged for 10 minutes at 1500 g and 4ºC. The cell pellets were washed twice with 30 mL of ice-cold ddH_2O. The cell pellets were re-suspended with 1 mL of ddH_2O and transferred to a 2 mL vial consisting of 0.1 mm silica beads. The cells were then

lysed using a FastPrep-24 instrument for 40 seconds at 6 m/s. The cellular mixture was centrifuged for 10 minutes at 12,400 g and 4ºC and the supernatant was extracted. 10 µL of the supernatant was used to determine the protein concentration in the extracted metabolomics sample. Bio-Rad DC Protein Assay was used to obtain the total protein concentration for each sample using bovine serum albumin as standard. The supernatant was frozen in a dry ice ethanol bath and stored at -80^{o}C until analyzed by NMR.

Two groups of triplicate and independent *M. smegmatis* mc^{2}155 cultures were also grown using $^{13}C_6$-glucose as the carbon-13 source. The two groups are untreated mc^{2}155 and mc^{2}155 treated with 75 µg/mL DCS. The cell cultures were grown in 50 mL of minimal media containing 22 mM $^{13}C_6$-glucose, 500 mM ammonium chloride, and essential salts. The pH of the minimal media was adjusted to 7.2. The cultures were grown as described above until an OD_{600} of 0.6 was achieved. DCS was then added to the selected cultures to a concentration of 75 µg/mL and the bacteria were allowed to grow for 2 more hours. The metabolome was extracted as described above.

Prior to collecting the 2D ^{1}H-^{13}C HSQC spectra, all metabolomics samples are lyophilized and re-suspended in a 500 mM potassium phosphate buffer in 99.8% D_2O at a pH of 7.2 (uncorrected) containing 500 µM of unlabeled 3-(trimethylsilyl)-propionic-2,2,3,3-d_4 acid sodium salt (TMSP) as an internal standard. The volume used for each sample was normalized based on the relative protein concentration determined for each sample. The sample with the lowest protein concentration was dissolved in 650 µL of the NMR buffer, where the other samples were dissolved in a proportionally larger buffer volume based on the relative protein concentrations. In this manner, the final

metabolomics sample concentrations were equivalent and directly comparable. 600 μL of the reconstituted metabolomics sample was then placed into a 5 mm NMR tube for data collection.

10.2.3 2D ^{1}H-^{13}C HSQC NMR Data Collection

The 2D ^{1}H-^{13}C HSQC NMR spectra were collected on a Bruker 500 MHz Avance DRX spectrometer equipped with a triple-resonance, Z-axis gradient cryoprobe. A BACS-120 sample changer with Bruker Icon software was used to automate the NMR data collection. All spectra were collected using the $HSQC_0$ pulse sequence and followed the experimental protocol as previously described [21]. The 2D ^{1}H-^{13}C HSQC NMR spectra were collected with a total of 2048 data points and a spectrum width of 5000.0 Hz, and 64 data points with a spectrum width of 17607.23 Hz in the ^{1}H and ^{13}C dimensions, respectively. A total of 16 dummy scans, 128 scans, a receiver gain of 9195.2, and a relaxation delay of 1.5 seconds was used to obtain all spectra.

10.2.4 2D ^{1}H-^{13}C HSQC NMR Data Analysis

All 2D NMR spectra were processed using the NMRpipe software package [22]. A reference spectrum for peak picking was created by adding all spectra together. All spectra including the reference spectra was automatically peak picked and the peak intensities were organize by their chemical shifts using NMRviewJ [23]. The observed NMR peaks were assigned to specific metabolites using ^{1}H and ^{13}C chemical shift tolerances of 0.05 and 0.40 ppm respectively. Metabominer [24], Madison Metabolomics

Consortium Database (MMCD) [25], Human Metabolome Database (HMDB) [26], and Platform for Riken Metabolomics (PRIME) [27, 28] were used to identify all metabolites. All identified metabolites were verified using KEGG [29] and MetaCyc [30] databases.

The concentrations of all metabolites were calculated using an extrapolation procedure that has been previously described [21]. The concentration for each metabolite from the triplicate datasets are then averaged, standard deviation was calculated, and Student's t-tests were used to test for statistical significant ($p < 0.05$) differences between each group.

10.2.5 NMR ligand binding assay for D-alanyl-D-alanine Ligase

M. tuberculosis D-alanyl-D-alanine ligase was obtained using protocols described previously [17]. Ligand binding assays were performed to determine if alanine, ATP, and DCS can bind to the ligase individually or in combination. Six combinations of each ligand were used: 1) ATP, 2) D-alanine 3) DCS, 4) ATP and D-alanine 5) ATP and DCS, and 6) ATP, D-alanine, and DCS. The final concentration for each ligand is 100 μM. A second set of the six combinations of ligands were prepared identically, but with the addition of 25 μM of D-alanyl-D-alanine ligase. NMR inhibition studies of the D-alanyl-D-alanine ligase were performed using multiple concentrations of D-alanine and DCS. A total of 4 mixtures were prepared using 100 μM D-alanine with 0, 250, 500, or 1000 μM DCS. ATP was kept at a high concentration of 6 mM to prevent any competitive inhibition by low ATP:ADP ratios. D-alanyl-D-alanine ligase was titrated into the solution for a final concentration of 25 μM.

Each of the NMR samples was dissolved in 600 μL of 50 mM Tris buffer (pH 8.0, uncorrected) consisting of 10 mM magnesium chloride, 11.1 μM TMSP, and 2% DMSO. 1D ^{1}H NMR spectra were collected using excitation sculpting to efficiently remove the solvent signal [31]. A total of 16k data points with a spectrum width of 5482.5 Hz were collected using 32 scans and 8 dummy scans. The 1D ^{1}H NMR spectra were processed using ACD 1D NMR manager version 12.0. The peak area for D-alanine and D-alanyl-D-alanine was determined and the concentration was calculated based on the TMSP peak.

10.3 Results

10.3.1 Overall Impact of DCS on Mycobacterium smegmatis and Mycobacterium tuberculosis Metabolomes

In our previous studies, we found that *M. smegmatis alr* null mutants were able to grow on Middlebrook 7H9 medium without D-alanine supplementation [32, 33]. Furthermore, principal component analysis (PCA) of NMR metabolomics data revealed that *M. smegmatis alr* null mutants had different clustering patterns in PCA scores plots than the wild type and resistant strains indicating that Alr inactivation had a major impact on the metabolome [9]. However, the wild type and *alr* mutant cells treated with DCS cluster together in the PCA scores plot, but separate from untreated *alr* mutant cells, indicating that DCS acts on a common target different from Alr. Also, the one-dimensional (1D) ^{1}H NMR metabolomics data showed that glutamate and pyruvate may be a source for D-alanine synthesis, suggesting that a transaminase may convert D-(L-)glutamate into D-alanine. This heterologous racemization of amino acids by a bacterial

transaminase has been previously observed [34]. Thus, Alr is not an essential function of mycobacteria and is not the lethal target of DCS. We hypothesized that the inhibition of a secondary target, possibly Ddl, may negatively impact cell survival and function as the lethal target.

To analyze in depth the mechanism of action of DCS in both *M. smegmatis* and *M. tuberculosis*, a comparison of the impact of DCS on both microorganisms was performed using NMR metabolomics (Figure 10.1). Typically, metabolites from cell extracts are detectable by ^{1}H-NMR where the intensities of each peak correlate with the concentrations of the metabolites. Therefore, any differences in peak intensity between organisms indicate a difference in the metabolomes. A spectral comparison between these two species indicates that there are few differences in the metabolomes (Figure 10.1a). More noticeable, the *M. tuberculosis* metabolome shows a higher concentration of carbohydrates and aminosugars (3.5-4.5 ppm). However, the impact of DCS is similar for both *M. smegmatis* and *M. tuberculosis* suggesting the mechanism of DCS inhibition is similar in both species. The 1D ^{1}H-NMR spectrum indicates that alanine (1.43-1.45 ppm) showed a large increase while glutamate (2.00-2.45 ppm) had a dramatic decrease in concentration when both *M. smegmatis* and *M. tuberculosis* were treated with DCS (highlighted by blue boxes). In *M. tuberculosis*, there is an increase in the concentration of UDP, acetate, α-ketoglutarate while there is a decrease in AMP, glutamine, and methionine (Figure. 10.1b). An increase in UDP would be expected because it is an important precursor for peptidoglycan synthesis by providing N-acetyl(glycolyl)-glucosamine and the corresponding muramate UDP-derivatives.

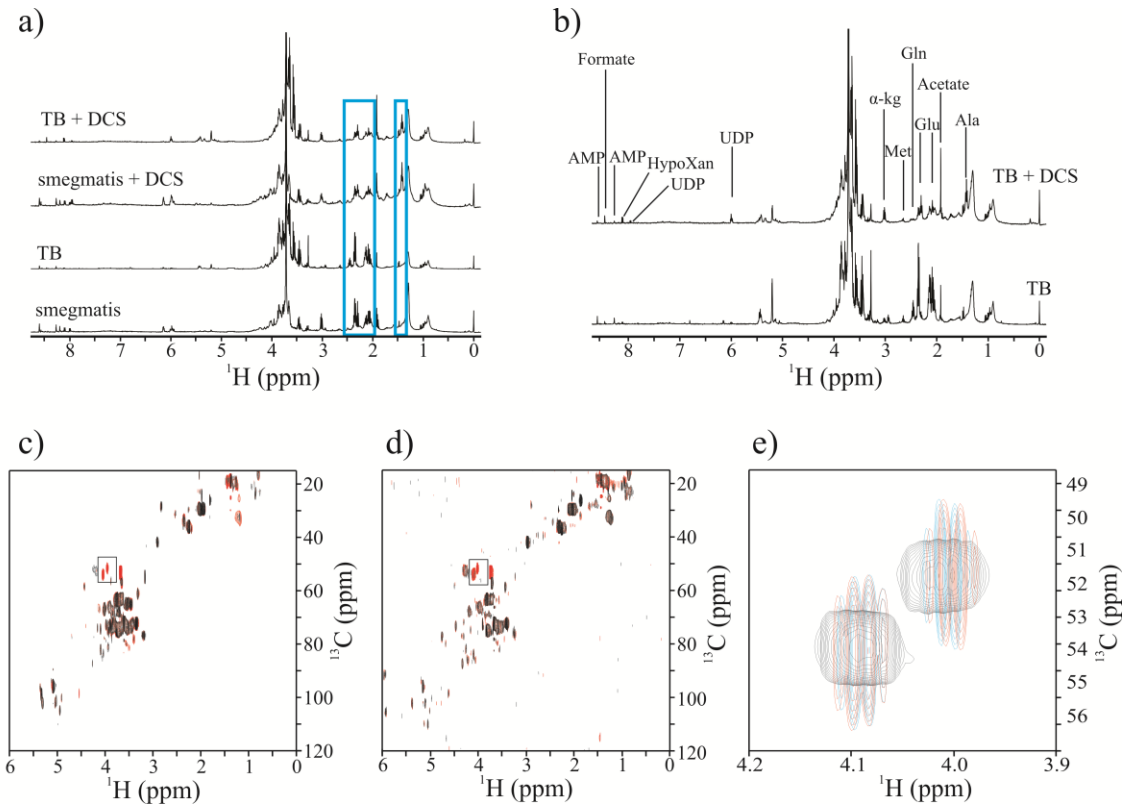


Figure 10.1. a) 1D ^{1}H NMR spectra comparing the metabolome of *M. tuberculosis* and *M. smegmatis* with (*top*) and without (*bottom*) treatment with DCS. The highlighted regions show the major differences between the extracted metabolomes when treated with DCS. **b)** The 1D ^{1}H NMR spectra of *M. tuberculosis* with (*top*) and without (*bottom*) treatment with DCS. Key metabolite changes are labeled. Overlay of 2D ^{1}H-^{13}C HSQC spectra comparing metabolite extracts from **c)** *M. tuberculosis* (black) and *M. tuberculosis* treated with 50 µg/mL DCS (red), **d)** *M. smegmatis* (black) and *M. smegmatis* treated with 75 µg/mL DCS (red). The circled region highlights the major differences between the untreated and treated cultures. **e)** Highlighted region of the 2D ^{1}H-^{13}C HSQC spectra comparing *M. tuberculosis* (red) and *M. smegmatis* (blue) treated with DCS, with a reference spectrum of D-alanyl-D-alanine (black).

To further quantify metabolite changes resulting from DCS treatment, we collected 2D ^{1}H-^{13}C HSQC spectra after supplementing the bacterial growth medium with a carbon-13 source. This procedure allows for monitoring the carbon-13 flow through the metabolome. Since the natural abundance of carbon-13 is 1.1%, only compounds derived from the ^{13}C-labeled metabolite through enzymatic turnover will bedetected in the 2D ^{1}H-^{13}C HSQC spectrum. As the D-alanine pathway is the primary focus of our study, both *M. smegmatis* and *M. tuberculosis* cells were pulse labeled with ^{13}C-D-alanine and treated with DCS shortly thereafter. Cells were further grown for approximately half to one generation and harvested. Not surprisingly, the resulting 2D ^{1}H-^{13}C HSQC spectrum of the metabolomes obtained from the two untreated mycobacterial species cells differed (Figure 10.1 c,d). As observed in the 1D spectra, the *M. tuberculosis* peaks in the 2D ^{1}H-^{13}C HSQC spectrum appear to be more congested in the carbohydrate and aminosugar region (^{13}C = 60-80 ppm, ^{1}H = 3.0-4.5 ppm). Therefore, the carbon-13 flow for *M. tuberculosis* appears to be directed towards gluconeogenesis. This differs from *M. smegmatis*, where the peaks are well dispersed throughout the 2D ^{1}H-^{13}C HSQC HSQC spectrum and correspondingly in the metabolome.

Again, the impact of DCS appears to be similar for both *M. smegmatis* and *M. tuberculosis* as three additional peaks clearly appeared in both 2D ^{1}H-^{13}C HSQC spectra upon drug treatment (circled region in Figure 10.1c,d). Using metabolomics databases containing reference NMR spectra, we were able to identify one of the peaks as alanine (^{13}Cα 56.0 ppm, ^{1}Hα 3.54 ppm). This is consistent with the inhibitory effect of DCS on Ddl that may lead to an accumulation of ^{13}C-D-alanine. However, the two nearby peaks

were not identified from the NMR databases. Based on the peak intensities and chemical shifts, it was suspected to be a close derivative of D-alanine. To test this hypothesis, we collected the natural abundance spectrum of a solution of D-alanyl-D-alanine as a reference. The two unknown peaks were a perfect match to this D-alanyl-D-alanine reference spectrum (Figure 10.1e) indicating that D-alanyl-D-alanine is accumulated upon DCS treatment with supplemental ^{13}C-D-alanine. This result was expected from Alr inhibition by DCS and the flow of the ^{13}C label into D-alanyl-D-alanine [35] as Ddl activity is maintained because the supplemented ^{13}C-D-alanine effectively competes with DCS for the Ddl active site [19]. Further inspection of cell growth curves confirmed that the DCS treated cultures when supplemented with ^{13}C-D-alanine were growing similarly to untreated cells. Thus, ^{13}C-D-alanine appears to compete very effectively with DCS and promote cell viability in the presence of DCS. This effect would readily explain the well-known reversal of DCS inhibition by D-alanine [36].

10.3.2 Impact of DCS on the Central Metabolism and Peptidoglycan Synthesis

To study the impact of DCS on *M. smegmatis*, a detailed analysis of changes to the metabolome was performed by measuring the ^{13}C metabolite concentrations using 2D time-zero ^{1}H-^{13}C HSQC ($HSQC_0$) experiments [21]. Two sets of triplicate cultures were grown in minimal media with ^{13}C-glucose as the primary carbon source. Cells were grown to exponential phase, treated with 75 μg/mL of DCS and grown for one additional generation. Analysis of the corresponding cell extracts (Figure 10.2a) indicated that the variations in the metabolites identified from the $HSQC_0$ experiment were similar to those

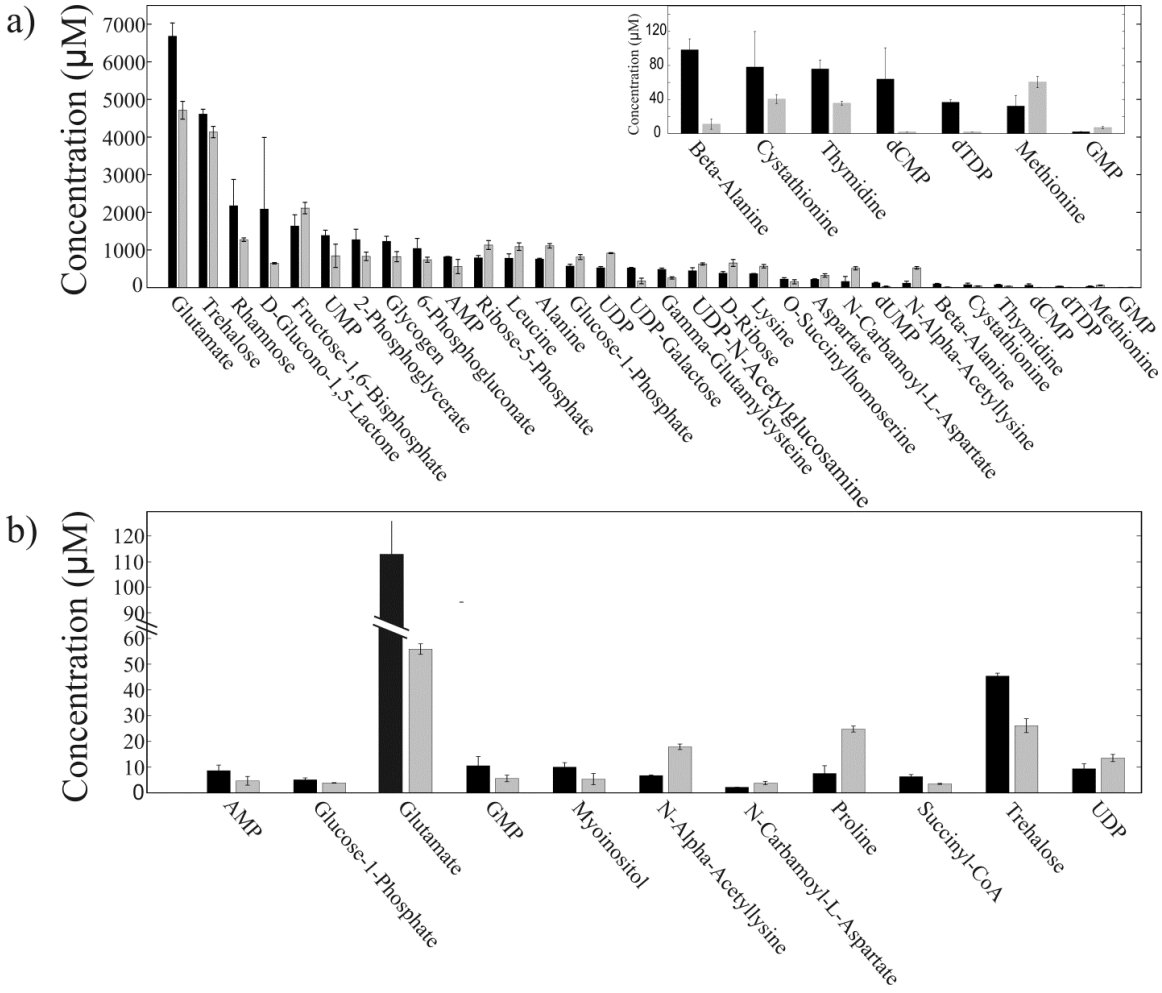


Figure 10.2. Bar graphs generated from the average of triplicate 2D ^{1}H-^{13}C HSQC spectra comparing the differences between untreated (black) and DCS treated (white) *M. smegmatis* cultures using **a)** 22 mM ^{13}C-glucose or **b)** 100 μM ^{13}C-pyruvate as the sole carbon-13 source. Metabolites having statistically significant perturbations ($p < 0.05$) upon treatment with DCS are displayed in the bar graph.

identified from the 1D ^{1}H-NMR spectra (Figure 10.1a,b). This further confirms that carbohydrate precursors leading towards the biosynthesis of trehalose, glycogen, rhamose and UDP-galactose were significantly decreased upon DCS treatment. Moreover, the increase in fructose-1,6-bisphosphate highlights the switch to the catabolic route suggesting up-regulation of the glycolytic pathway. Glutamate is an important metabolite for energy generation, purine synthesis, peptidoglycan synthesis, and transfer of amino groups. Consistent with these roles, glutamate was greatly decreased in the presence of DCS suggesting the metabolite was catabolized. This same catabolic effect was observed on the pools of nucleotide precursors such as dUMP, dCMP, dTDP, and thymidine, which were all decreased significantly upon DCS treatment. There is also a significant decrease in the concentration of precursors in the oxidative branch of the pentose phosphate pathway such as D-glucono-1,5-lactone and 6-phosphogluconate. But, there is also an increase in the concentration of D-ribose-5-phosphate and D-ribose, suggesting a switch to the non-oxidative branch of the pentose phosphate pathway.

β-alanine is the precursor of important vitamins such as panthotenate via the conversion of aspartate to β-alanine by the PanD enzyme. Moreover, pantothenate is a precursor of Coenzyme A, which is essential for the production of fatty acids and peptidoglycan. A significant decrease in the pool of β-alanine was also observed, potentially suggesting DCS may inhibit the PanD enzyme. Instead, the carbon flow is rerouted towards the production of methionine and lysine, the terminal product of the pantothenate pathway immediately downstream from meso-2-6-diaminopimelate, a component of the peptidoglycan bridge. However, the increase in concentration of N-

alpha-acetyl-lysine suggests that lysine is also being catabolized. Other peptidoglycan precursors such as alanine, UDP-N-acetyl-glucosamine, UDP, D-ribose, D-ribose-5-phosphate, and carbamoyl-L-aspartate are all significantly increased, consistent with the inhibition of peptidoglycan biosynthesis.

These results were confirmed by a second metabolomics study using ^{13}C-pyruvate as a carbon-13 source (Figure 10.2b). Carbon flow was initially directed toward gluconeogenesis and glutamate production. However, trehalose, myoinositol, glucose-1-phosphate and glutamate decreased significantly upon DCS treatment. Consistent with the inhibition of peptidoglycan synthesis, the carbon flow was also directed towards purine and lysine biosynthesis, as observed by the significant increase in UDP, N-carbamoyl-L-aspartate, and N-alpha-acetyl-lysine. There was also a significant increase in proline concentration as the metabolism of this amino acid plays an important role in the detoxification of methylglyoxal, a toxic aldehyde generated from the cleavage of the phosphate group of dihydroxyactenone phosphate by the enzyme methylglyoxal synthase [37]. The over production of methylglyoxal has been associated with the imbalance in the anabolic and catabolic processes in the cell [38]. Although we were unable to identify methylglyoxal, the increase in proline levels suggests the existence of a mechanism to compensate for excessive activity in the catabolic pathways.

10.3.3 The effect of D-alanine on DCS inhibitory activity

To determine the effect of DCS on the D-alanine pathway, *M. smegmatis* cultures were grown in Middlebrook 7H9 media supplemented with 100 μM ^{13}C-D-alanine at

mid-exponential phase. After 10 min incubation, cultures were treated with either 75, 300, or 1200 μg/mL of DCS. To determine the uptake of ^{13}C-D-alanine, the total area under the peaks observed in the 2D ^{1}H-^{13}C HSQC spectra was compared between treated and untreated samples. This is possible because all of the peaks in the 2D ^{1}H-^{13}C HSQC spectrum are derived from ^{13}C-D-alanine. We observed that D-alanine was able to effectively compete with DCS for uptake. At 75 and 300 μg/mL the total carbon-13 concentration was approximately 35% higher in the DCS-treated cultures. This indicated that D-alanine uptake was increased as needed to produce peptidoglycan precursors and increase the internal D-alanine pools to out-compete DCS from internal targets such as Alr and Ddl. However, a 15% decrease in D-alanine uptake was observed at 1200 μg/mL DCS, which represents a 100-fold molar excess of DCS.

D-alanine readily reversed cell growth inhibition by DCS. For example, cell growth was inhibited when cultures were grown with ^{13}C-pyruvate and ^{13}C-glucose using 75 μg/mL DCS. Upon addition of 100 μM ^{13}C-D-alanine to the culture medium, growth inhibition was not observed even with 300 μg/mL DCS. Consistent cell growth inhibition only occurred when the DCS dosage was raised to 1200 μg/mL, a 100-fold molar excess of DCS to D-alanine (Figure 10.3a). Comparison between non-inhibitory and inhibitory conditions was important to assess the transient effect of DCS on the metabolome (Figure 10.3b). At 75 or 300 μg/mL DCS, the overall impact on the metabolome was minimal, as indicated by the corresponding heat maps of metabolite concentrations derived from ^{13}C-D-alanine (Figure 10.3a). However, major changes were observed in cultures treated with

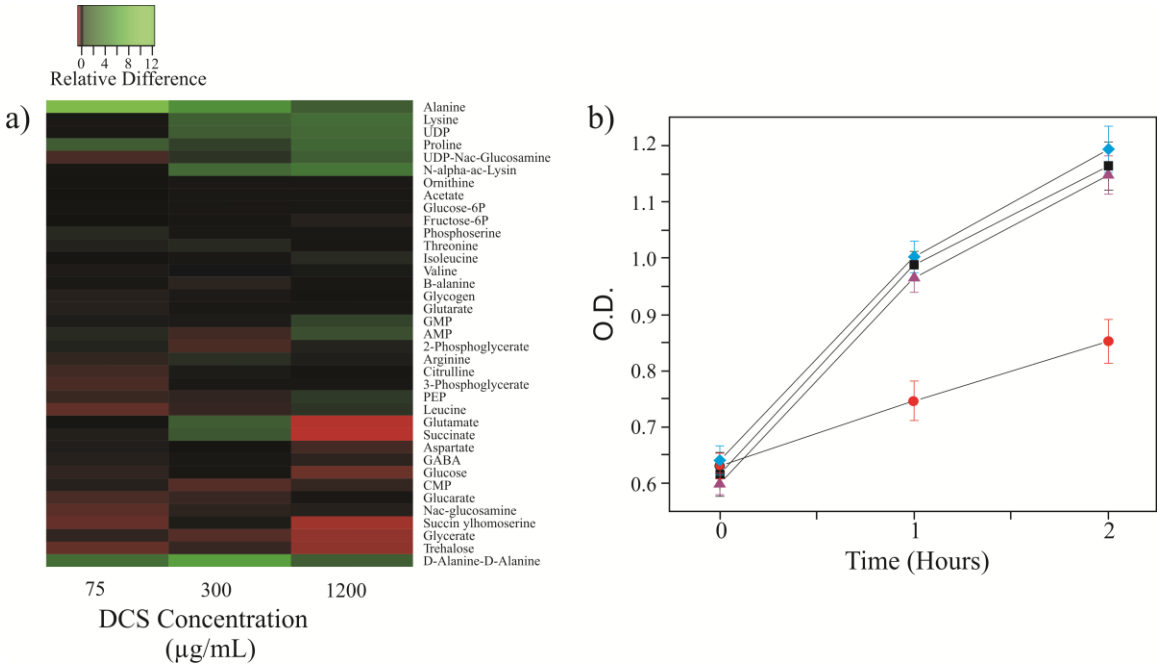


Figure 10.3. a) Heat map generated from the average of triplicate $HSQC_0$ spectra comparing the relative difference in metabolite concentrations between *M. smegmatis* and *M. smegmatis* treated with 75, 300, or 1200 µg/mL DCS using 100 µM ^{13}C-D-alanine as the carbon-13 source. The relative differences between the untreated and treated cultures are plotted on a color scale from -100% to -25% (red), -25% to 25% (black), and 25% to 1,200% (green). **b)** *M. smegmatis* growth curve comparing untreated (black) cultures to cultures treated with 75 (magenta), 300 (blue), 1200 µg/mL (red) DCS. O.D values were determined after the treatment of DCS with an initial $O.D._{600}$ of ~0.6.

1200 μg/mL DCS, a concentration resulting in inhibition of both D-alanine uptake and cell growth.

Increasing the DCS concentration had a major impact on the pool of ^{13}C-labeled metabolites originating from ^{13}C-D-alanine (Figure 10.4). At 75 μg/mL DCS, the internal pool of D-alanine increased 12-fold with respect to untreated culture. Similarly, a 4-fold increase in D-alanyl-D-alanine and a slight 10% increase in glutamate were also observed. At 300 μg/mL DCS, the D-alanine concentration increased 4-fold compared to untreated cultures, while the D-alanyl-D-alanine and glutamate concentrations increased further by 5.5-fold and 50%, respectively. At 1200 μg/mL, both the D-alanine and D-alanyl-D-alanine concentrations were similar in both treated and untreated cultures, while the glutamate concentration decreased by 50%. Treatment with increasing DCS concentrations also leads to similar changes in other metabolite pools. For example, an increase in DCS concentration results in an increase in peptidoglycan precursors such as lysine and UDP. These results are consistent with the hypothesis that D-alanine can be metabolized by three pathways: i) direct stereospecific conversion into L-alanine by Alr, ii) dimerization into the peptidoglycan precursor D-alanyl-D-alanine by Ddl, and iii) transamination into pyruvate with concomitant formation of glutamate from α-ketoglutarate (Figure 10.5). As pyruvate turnover is extremely fast (Figure 10.2b), the rapid incorporation of the ^{13}C-label into glutamate may proceed in three steps: i) conversion of D-alanine into pyruvate, ii) pyruvate turnover into α-ketoglutarate, and finally iii) the incorporation of the ^{13}C-label into glutamate by iteration of the first step as required by the principle of microscopic reversibility. Thus, in absence of DCS, D-

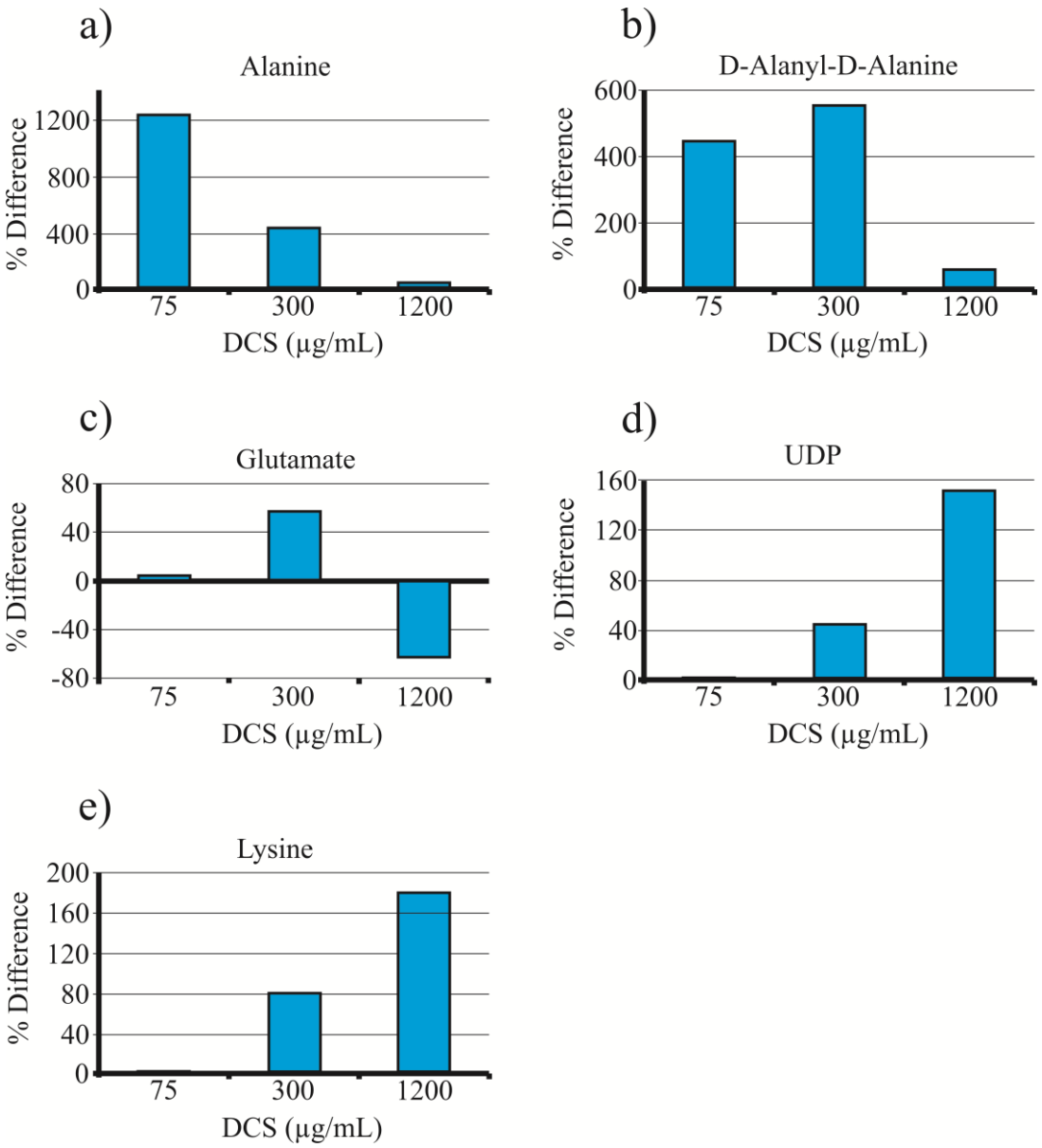


Figure 10.4. Bar graphs depicting relative concentration differences in metabolite precursors **a)** alanine, **b)** D-alanyl-D-alanine, UDP, **c)** glutamate, **d)**UDP, and **e)** lysine that are involved in the biosynthesis of peptidoglycan. The observed metabolite concentration changes result from treating *M. smegmatis* and *M. smegmatis* with DCS. A positive value indicates an increase in the concentration of the metabolites when the cultures are treated with DCS, and a negative value indicates a decrease in the concentration of the metabolites when the cultures are treated with DCS.

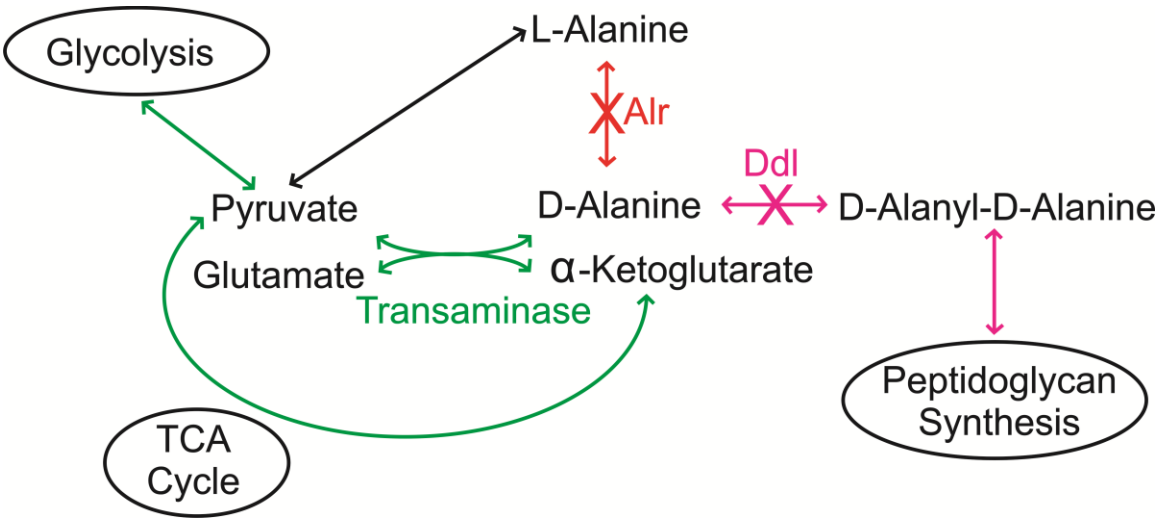


Figure 10.5. Schematic representation of pathways that may synthesize D-alanine. Alr (red path) is inhibited at a dose of 75 μg/mL of DCS, which inhibits the primary pathway of D-alanine production. However, inhibiting Alr is non-lethal as the transaminase pathway (green path) is hypothesized to remain unaffected upon DCS treatment. This reaction may lead to an alternative mechanism of D-alanine biosynthesis that antagonizes DCS activity through a competition for Ddl. Ddl (pink path) is inhibited at a dose of 1200 μg/mL of DCS, inhibition of peptidoglycan biosynthesis initiates a cascade of events leading to cell death [39] TCA, tricarboxylic acid.

alanine is rapidly converted into L-alanine by Alr as well as other metabolites such as D-alanyl-D-alanine and glutamate. As the concentration of DCS is increased, Alr inhibition leads to label accumulation in the alanine pool, while the D-alanyl-D-alanine pool remains constant only decreasing at the higher DCS concentrations (1200 μg/mL) reflecting the inhibition of Ddl. Moreover, though the conversion of D-alanine into L-alanine is inhibited, the results indicate that part of this label is still converted into glutamate. These results suggest that DCS first inhibits Alr, but cell growth inhibition parallels the inhibitory effect of DCS on Ddl. Moreover, there is a conversion of D-alanine into glutamate at all DCS concentrations, suggesting that a DCS-insensitive D-alanine transaminase converts D-alanine into glutamate by steps described above. By the principle of microscopic reversibility, this enzyme may also convert glutamate into D-alanine. Nonetheless, the major conversion of alanine into glutamate depends on L-alanine transaminase, as the ^{13}C-label incorporated into glutamate decreases significantly upon Alr inhibition. Therefore, Ddl rather than Alr, is the main lethal target of DCS, while D-alanine transaminase is responsible for the alternative pathway of D-alanine biosynthesis (Figure 10.5). It is important to note that other enzymatic reactions that produce D-alanine are possible alternative explanations, such as the Strickland reaction in the ornithine fermentation pathway [40] or the decarboxylation of D-(L-)-aspartate by a D-amino transaminase [41]. Correspondingly, the observed PanD inhibition and decrease in β-alanine caused by DCS may result from an increase in the production of CoA for the Strickland reaction. Of course, the ornithine fermentation pathway that is composed of nine enzymes has not been identified in mycobacteria. Alternatively, the observed

increase in aspartate upon DCS treatment may suggest D-aspartate is being converted into D-alanine through the activity of a D-amino transaminase. Nevertheless, this still requires invoking the activity of a transaminase to explain the production of D-alanine when Alr is inhibited by DCS.

10.3.4 Impact of DCS on D-alanyl-D-alanine Ligase

To further investigate the interaction of DCS and the Ddl lethal target, the binding of relevant potential ligands with Ddl were followed by NMR. Moreover, the influence of a given ligand on the simultaneous or consecutive binding of another ligand was followed as well. Previous studies have yielded dissimilar results regarding the dependency of D-alanine binding on the presence of ATP in solution [8, 17]. To resolve this issue, purified Ddl was incubated with ATP, D-alanine, or both ATP and D-alanine and binding was analyzed by NMR (Figure 10.6-10.8). The observed NMR peak broadening for ATP upon the addition of Ddl indicates that ATP binds strongly to Ddl in the absence of D-alanine. Conversely, the D-alanine NMR peaked is unaffected by the addition of Ddl in the absence of ATP, indicating that D-alanine cannot bind to Ddl without ATP. The addition of ATP induces the binding of D-alanine to Ddl and the conversion of D-alanine to D-alanyl-D-alanine and ATP to ADP (Figure 10.9). These results are consistent with the mechanism of ordered binding and the binding assays reported by Prosser [8]. Ddl NMR line-broadening binding assays were also carried out with DCS, DCS and ATP, and DCS with both ATP and D-alanine. The DCS NMR peaks broaden and incurred a chemical shift change in all three cases indicating that, in contrast to D-alanine, DCS can

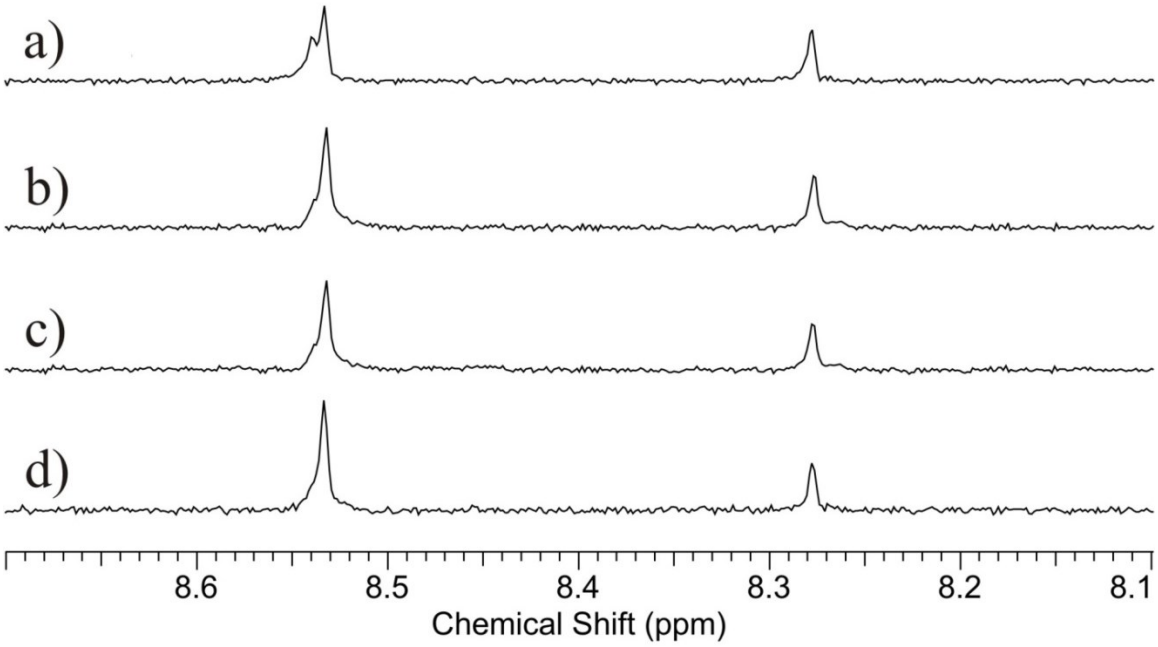


Figure 10.6. Ligand binding assays comparing the intensities of the ATP peaks (8.28 and 8.54ppm) in mixtures consisting of: a) 100 µM ATP, D-alanine, DCS and 25 µM Ddl; b) 100 µM ATP and D-alanine and 25 µM Ddl; c) 100 µM ATP and 25 µM Ddl; and d) 100 µM ATP as a control.

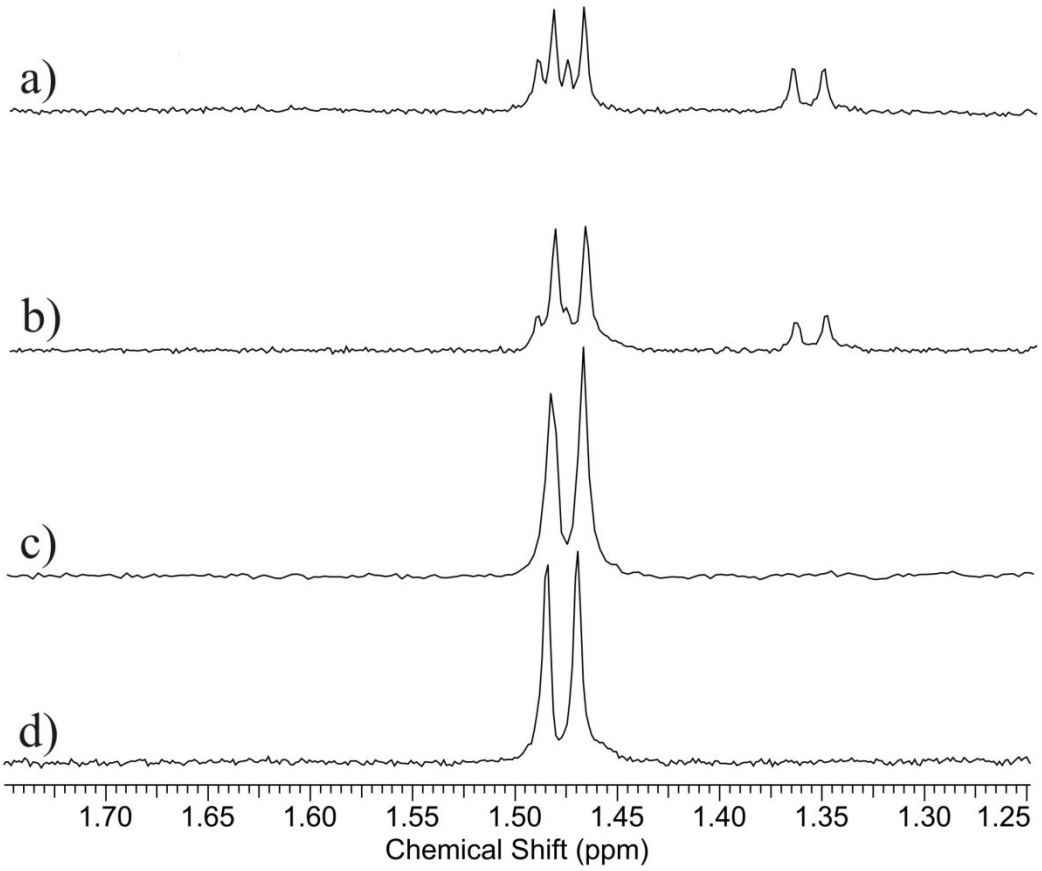


Figure 10.7. Ligand binding assays comparing the intensities of the D-alanine peaks (1.45-1.50) in mixtures consisting of: a) 100 μM ATP, D-alanine, DCS and 25 μM Ddl; b) 100 μM ATP and D-alanine and 25 μM Ddl; c) 100 μM D-alanine and 25 μM Ddl; and d) 100 μM D-alanine as a control. Formation of D-alanyl-D-alanine (1.33-1.38ppm) is formed when ATP, D-alanine, and Ddl are in the same mixture.

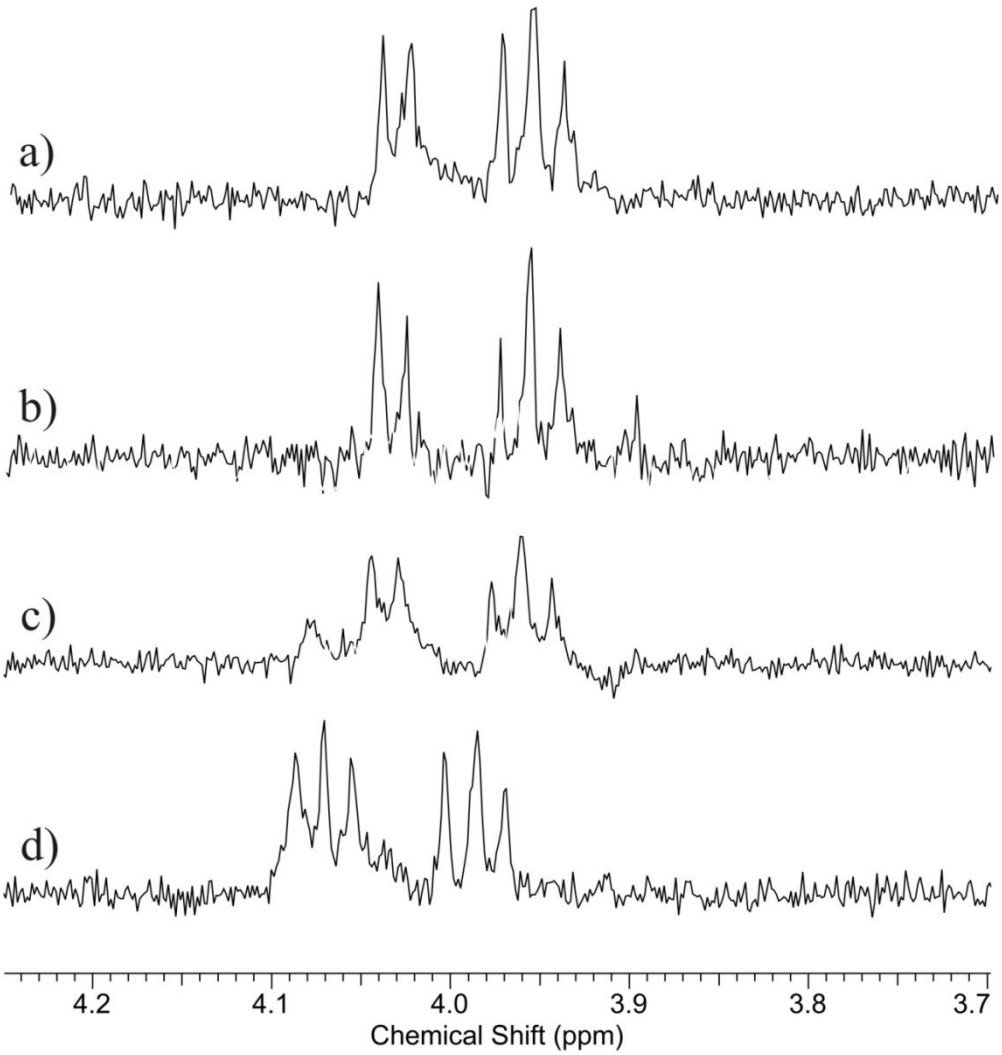


Figure 10.8. Ligand binding assays comparing the intensities of the DCS peaks (3.95-4.10 ppm) in mixtures consisting of: a) 100 μM ATP, D-alanine, DCS and 25 μM Ddl; b) 100 μM ATP, DCS and 25 μM Ddl; c) 100 μM DCS and 25 μM Ddl; and d) 100 μM DCS as a control.

bind Ddl independently of ATP. Nonetheless, DCS is a weak binding ligand as evident by the modest changes in line-width and chemical shifts consistent with reported K_i values of 14 μM[8] to 0.9 mM.[42]. Furthermore, no additional peaks were observed suggesting that there was no chemical modification to DCS. This suggests DCS is a competive inhibitor that binds to Ddl with no chemical modification. This observation is unique compared to inhibition to Alr and D-alanine transaminase. Alr and D-alanine transaminase are pyridoxal-phosphate (PLP) dependent enzymes that are also inhibited by DCS. Unexpectedly, DCS and PLP forms a complex stable aromatic tautomer which then inhibits the PLP dependent enzymes [43].

To confirm these results, we followed the progression of the Ddl (25 μM) reaction at varying DCS concentrations and fixed amounts of D-alanine (100 μM) and ATP (100 μM). Using NMR, the concentrations of the substrates (ATP and D-alanine) and two of the products (ADP and D-alanyl-D-alanine) were determined over time (Figure 10.9). In the absence of DCS, the conversion of ATP and D-alanine into ADP and D-alanyl-D-alanine was stoichiometric, but followed a very slow kinetics, reaching a maximum conversion in the forward reaction at approximately 4 h (Figure 10.10a). Thereafter, the reaction was inhibited by accumulation of the ADP and D-alanyl-D-alanine products. To analyze the inhibitory effect of DCS, the reaction conditions were modified by adding saturating amounts of ATP at 6.0 mM. Under these conditions, the reaction proceeds to completion until D-alanine is exhausted (Figure 10.10b). At increasing DCS concentrations, the reaction is progressively inhibited, but never reaches full inhibition.

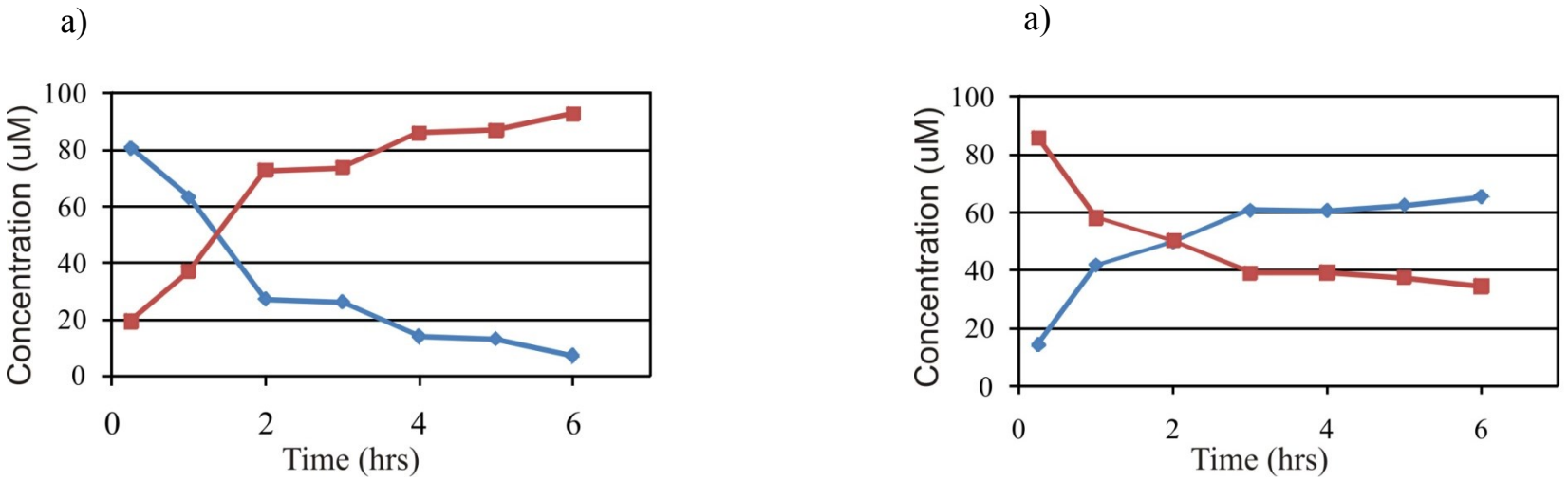


Figure 10.9. 1D ^{1}H NMR was used to monitor the conversion of **a)** ATP (blue) to ADP (red) and **b)** D-alanine (red) to D-alanyl-D-alanine (blue) by Ddl. The initial concentrations were fixed at Ddl (25 μM), D-alanine (100 μM), and ATP (100 μM).

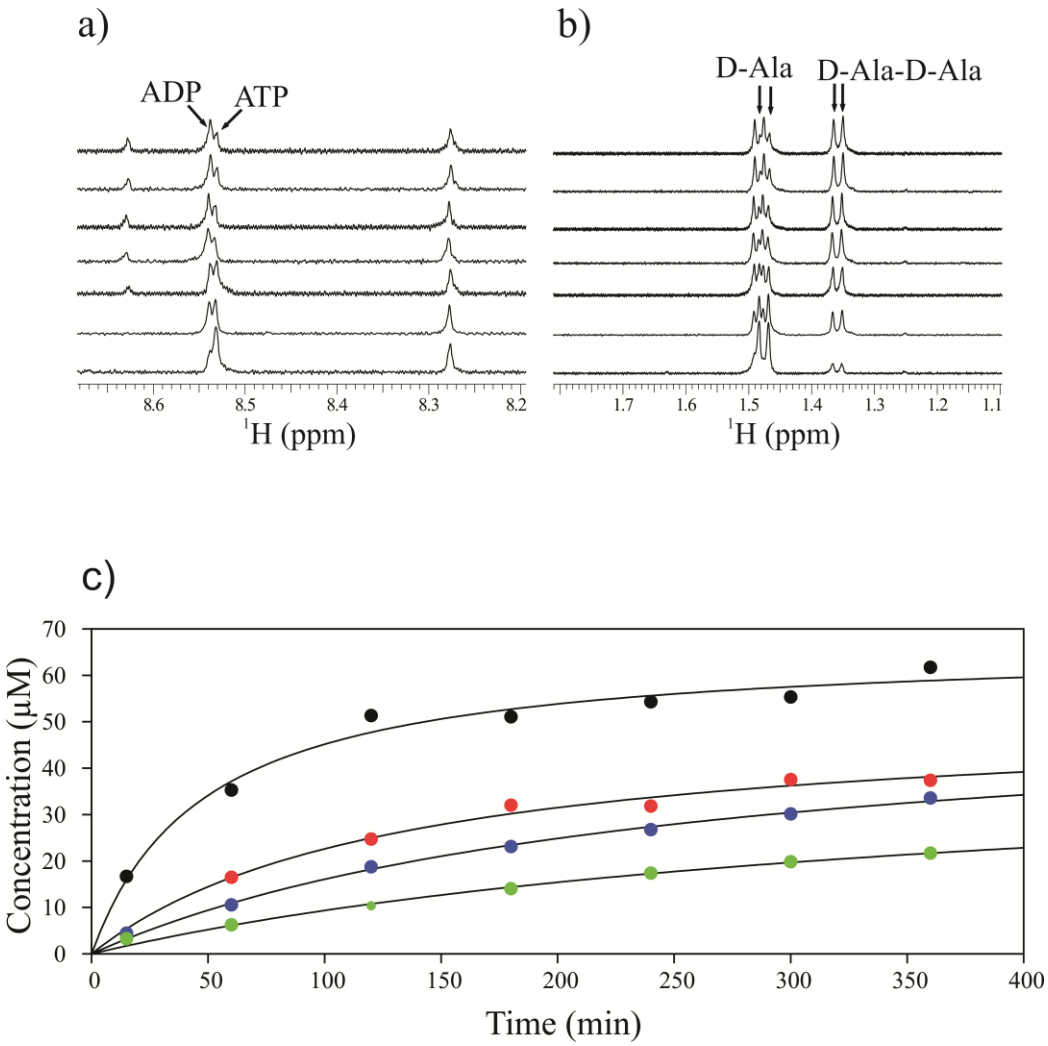


Figure 10.10. NMR kinetic analysis of **a)** ADP and **b)** D-alanyl-D-alanine formation by Ddl. The 1D ^{1}H spectrum were collected at one hour intervals after the addition of 25 μM of Ddl into a mixture containing 100 μM D-alanine and 100 μM ATP. **c)** Plot of D-alanyl-D-alanine formation as monitored by 1D ^{1}H NMR that demonstrates a decrease in Ddl activity as a function of increasing DCS concentration: 0 (black), 250 (red), 500 (blue), 1000 (green) μM of DCS.

This is as expected given that DCS binds weakly to Ddl and D-alanine is a competitor of DCS.

10.4 Discussion

DCS is known to inhibit peptidoglycan biosynthesis, but its primary target has been a point of controversy and extensive investigation for over fifty years. The effect of DCS has been attributed to the inhibition of Alr, Ddl or both [32, 33, 44, 45]. Both Alr and Ddl are known binding targets in mycobacteria and their enzymatic activities are inhibited by DCS in a concentration dependent manner [19, 35]. Alr is the only known enzyme to produce D-alanine in *M. smegmatis* and *M. tuberuculosis.* In addition,previous studies have shown that the overexpression of *alr* confers resistance to DCS while *alr* mutant strains are more susceptible [19, 33]. However, overexpression of *ddl* also leads to DCS resistance, but to lower levels [35]. Similarly, contradictory genetic studies have been presented that seem to indicate that Alr is required in the absence of D-alanine [44, 45]. Nonetheless, as we have previously observed [19] these seemingly contradictory results may occur because of the different experimental conditions employed by these various studies to analyze Alr conditional essentiality. Critically, *alr* mutants are still able to grow in the absence of D-alanine [32, 33]. Also, the NMR metabolomic profiles of *alr* mutants do not match the metabolome of wild type *M. smegmatis* treated with DCS, which indicates that Alr is not the main inhibitory target of DCS [9]. Thus, these studies taken together conclusively rule out Alr as the lethal target of DCS in mycobacteria.

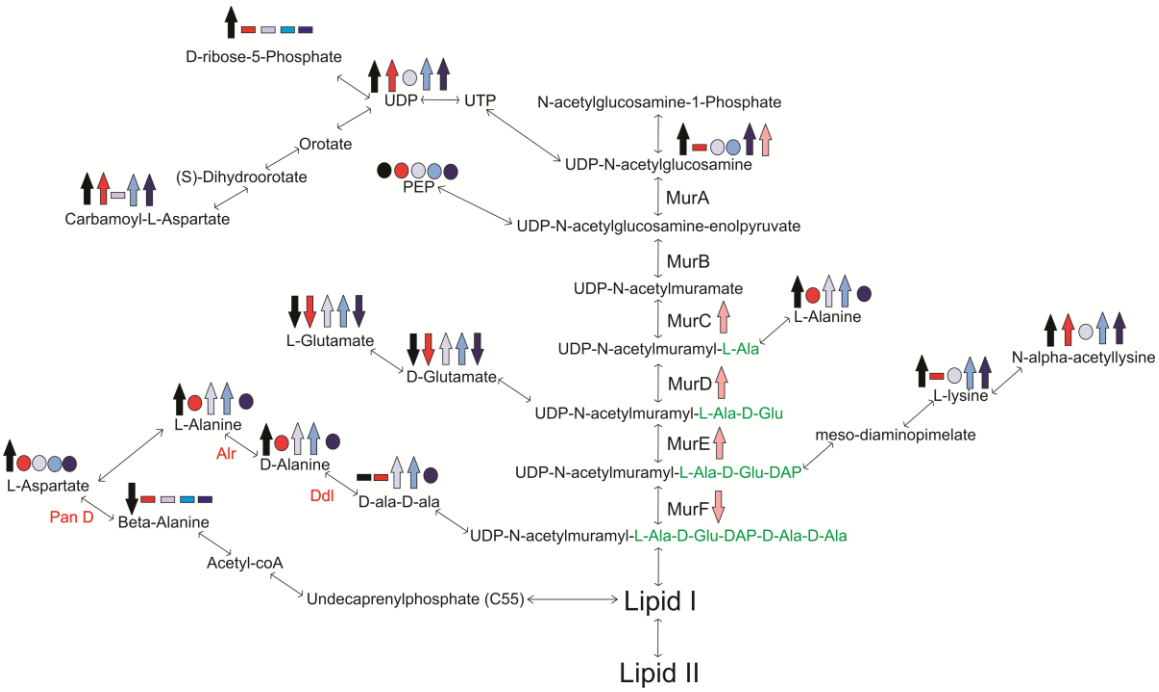


Figure 10.11. Peptidoglycan biosynthesis pathway depicting the metabolites that were identified by 2D ^{1}H-^{13}C HSQC analysis or from literature results.[46, 47] The arrows correspond to metabolites with statistically significant ($p < 0.05$) concentration increases (up) or decreases (down) when comparing DCS treated cells to untreated cells. Circles indicate the metabolite concentration is similar ($p > 0.05$). The p values were calculated using the Student's t-test. Hyphens indicate the metabolites were not identified for the specific carbon-13 source. Cells were incubated with ^{13}C-glucose and treated with 75 µg/mL of DCS (black arrow), ^{13}C-pyruvate and treated with 75 µg/mL of DCS (red arrow), ^{13}C-alanine and treated with 75 µg/mL of DCS (grey arrow), ^{13}C-alanine and treated with 300 µg/mL of DCS (blue arrow), ^{13}C-alanine and treated with 1200 µg/mL of DCS (dark blue arrow), and literature data (pink arrows)

In this study, we found that DCS does not only inhibit peptidoglycan synthesis, but it also caused a cascade of effects on the central metabolism of the cell (Figure 10.11). Our NMR metabolomics results using both ^{13}C-glucose and ^{13}C-pyruvate as carbon-13 sources showed a large increase in the accumulation of peptidoglycan precursors, as would be expected from DCS inhibition of peptidoglycan biosynthesis. Furthermore, DCS treatment of both *M. smegmatis* and *M. tuberculosis* resulted in a metabolic shift towards a catabolic state. We hypothesize that this catabolic shift occurs to compensate for the need to increase the synthesis of peptidoglycan precursors used to construct the cell wall. The shift towards the non-oxidative branch of the pentose phosphate pathway was evidenced by a decrease in oxidative branch metabolites with a concomitant increase in ribose-5-phosphate. This effect is most likely associated with the inhibition of peptidoglycan synthesis as ribose-5-phosphate is used in the synthesis of UTP needed to generate UDP-N-acetyl-glucosamine, which is the initial precursor in this biosynthetic step. This metabolic shift is also likely to result in a decrease in mycolic acid formation since the pentose phosphate oxidative branch is a major source of NADPH, which is required for the synthesis of mycolic and fatty acids. This would be expected to further weaken the cell wall and contribute to the bactericidal action of DCS.

The demonstration that a decrease in the biosynthesis of D-alanyl-D-alanine occurs simultaneously with cell growth inhibition shows that Ddl is the primary target of DCS. However, inhibition of Alr may contribute indirectly to the effect of DCS by simply lowering the pools of D-alanine produced [32]. The lower levels of D-alanine glutamate as we report here and elsewhere [9, 19, 32, 33]. Other investigators proposed

that Alr, rather than Ddl, is the main target of DCS based on the reversal of growth inhibition by externally added D-alanine and the weak reversal activity of D-alanyl-D-alanine [44]. However, our results clearly show that D-alanine and DCS are both competitors for Ddl binding, where a significant increase in D-alanine concentration can mitigate DCS activity by the intracellular protection of Ddl. Ddl is simply a kinetically alternative pathway by transamination from pyruvate with concomitant production of inefficient enzyme and, correspondingly, a bottle neck in peptidoglycan biosynthesis. The externally added D-alanine is simply increasing Ddl activity by outcompeting DCS.upon Alr inactivation may allow DCS to outcompete D-alanine for Ddl binding. However the inactivation of Alr would not prevent the cells from obtaining D-alanine from an Thus, increasing the intracellular pool of D-alanine is an effective mechanism of increasing DCS resistance. In conclusion, our combined analysis in this and previous studies indicate that DCS acts primarily on Ddl, an alternative pathway provides a source of D-alanine when Alr activity is inhibited or deleted, and Alr plays an indirect role in protecting Ddl by maintaining a higher internal pool of D-alanine.

10.5 References

1. World Health Organization, *Global Tuberculosis Report 2012*, 2012, World Health Organization: Geneva, Switzerland:. p. p. 100.

2. Zhang, Y., *The Magic Bullets and Tuberculosis Drug Targets.* Annual Review of Pharmacology and Toxicology, 2005. **45**: p. 529-64.

3. Mitnick, C.D., et al., *Comprehensive Treatment of Extensively Drug-Resistant Tuberculosis.* The New England Journal of Medicine, 2008. **359**(6): p. 563-74.

4. Cunha, B.A., *Antibiotic Side Effects.* The Medical Clinics of North America, 2001. **85**(1): p. 149-85.

5. Torun, T., et al., *Side Effects Associated With the Treatment of Multidrug-Resistant Tuberculosis.* The International Journal of Tuberculosis and Lung Disease : The Official Journal of the International Union Against Tuberculosis and Lung Disease, 2005. **9**(12): p. 1373-7.

6. Hood, W.F., R.P. Compton, and J.B. Monahan, *D-Cycloserine: A Ligand for the N-Methyl-D-Aspartate Coupled Glycine Receptor Has Partial Agonist Characteristics.* Neuroscience Letters, 1989. **98**(1): p. 91-5.

7. Lambert, M.P. and F.C. Neuhaus, *Mechanism of D-Cycloserine Action: Alanine Racemase From Escherichia coli W.* Journal of Bacteriology, 1972. **110**(3): p. 978-87.

8. Prosser, G.A. and L.P. de Carvalho, *Kinetic Mechanism and Inhibition of Mycobacterium tuberculosis D-Alanine:D-Alanine Ligase by the Antibiotic D-Cycloserine.* The FEBS journal, 2013. **280**(4): p. 1150-66.

9. Halouska, S., et al., *Use of NMR Metabolomics to Analyze the Targets of D-Cycloserine in Mycobacteria: Role of D-Alanine Racemase.* Journal of Proteome Research, 2007. **6**(12): p. 4608-14.

10. Takayama, K., et al., *Isolation and Characterization of Uridine Diphosphate-N-Glycolylmuramyl-L-Alanyl-Gamma-D-Glutamyl-Meso-Alpha,Alpha'-Diamino Pimelic Acid from Mycobacterium tuberculosis.* Biochemical and Biophysical Research Communications, 1970. **39**(1): p. 7-12.

11. Lovering, A.L., S.S. Safadi, and N.C. Strynadka, *Structural Perspective of Peptidoglycan Biosynthesis and Assembly.* Annual Review of Biochemistry, 2012. **81**: p. 451-78.

12. Crick, D.C., S. Mahapatra, and P.J. Brennan, *Biosynthesis of the Arabinogalactan-Peptidoglycan Complex of Mycobacterium tuberculosis.* Glycobiology, 2001. **11**(9): p. 107R-118R.

13. Vollmer, W., D. Blanot, and M.A. de Pedro, *Peptidoglycan Structure and Architecture.* FEMS Microbiology Reviews, 2008. **32**(2): p. 149-67.

14. de Souza, M.V.N., et al., *Synthesis and Biological Aspects of Mycolic Acids: An Important Target against Mycobacterium tuberculosis.* The Scientific World Journal, 2008. **8**: p. 720-751.

15. Strych, U., et al., *Characterization of the Alanine Racemases from Two Mycobacteria.* FEMS Microbiology Letters, 2001. **196**(2): p. 93-8.

16. LeMagueres, P., et al., *The 1.9 A Crystal Structure of Alanine Racemase From Mycobacterium tuberculosis Contains A Conserved Entryway Into The Active Site.* Biochemistry, 2005. **44**(5): p. 1471-81.

17. Bruning, J.B., et al., *Structure of the Mycobacterium tuberculosis D-Alanine:D-Alanine Ligase, a Target of the Antituberculosis Drug D-Cycloserine.* Antimicrobial Agents and Chemotherapy, 2011. **55**(1): p. 291-301.

18. El Zoeiby, A., F. Sanschagrin, and R.C. Levesque, *Structure and Function of the Mur Enzymes: Development of Novel Inhibitors.* Molecular Microbiology, 2003. **47**(1): p. 1-12.

19. Caceres, N.E., et al., *Overexpression of the D-Alanine Racemase Gene Confers Resistance to D-Cycloserine in Mycobacterium smegmatis.* Journal of Bacteriology, 1997. **179**(16): p. 5046-55.

20. Zinniel, D.K., et al., *Sample Preparation of Mycobacterium tuberculosis Extracts for Nuclear Magnetic Resonance (NMR) Metabolomic Studies.* Journal of Visualized Experiments, 2012. **67**: p. e3673.

21. Hu, K.F., W.M. Westler, and J.L. Markley, *Simultaneous Quantification and Identification of Individual Chemicals in Metabolite Mixtures by Two-Dimensional Extrapolated Time-Zero H-1-C-13 HSQC (HSQC(0)).* Journal of the American Chemical Society, 2011. **133**(6): p. 1662-1665.

22. Delaglio, F., et al., *NMRPipe: A Multidimensional Spectral Processing System Based on UNIX Pipes.* Journal of Biomolecular NMR, 1995. **6**(3): p. 277-93.

23. Johnson, B.A., *Using NMRView to Visualize and Analyze the NMR Spectra of Macromolecules.* Methods in Molecular Biology, 2004. **278**: p. 313-52.

24. Xia, J., et al., *MetaboMiner--Semi-Automated Identification of Metabolites from 2D NMR Spectra of Complex Biofluids.* BMC Bioinformatics, 2008. **9**: p. 507.

25. Cui, Q., et al., *Metabolite Identification via the Madison Metabolomics Consortium Database.* Nature Biotechnology, 2008. **26**(2): p. 162-4.

26. Wishart, D.S., et al., *HMDB 3.0--The Human Metabolome Database in 2013.* Nucleic Acids Research, 2013. **41**(Database issue): p. D801-7.

27. Sakurai, T., et al., *PRIMe Update: Innovative Content for Plant Metabolomics and Integration of Gene Expression and Metabolite Accumulation.* Plant & Cell Physiology, 2013. **54**(2): p. e5.

28. Akiyama, K., et al., *PRIMe: A Web Site that Assembles Tools for Metabolomics and Transcriptomics.* In silico biology, 2008. **8**(3-4): p. 339-45.

29. Kanehisa, M., *The KEGG Database.* Novartis Foundation symposium, 2002. **247**: p. 91-101; discussion 101-3, 119-28, 244-52.

30. Caspi, R., et al., *The MetaCyc Database of Metabolic Pathways and Enzymes and the BioCyc Collection of Pathway/Genome Databases.* Nucleic Acids Research, 2008. **36**(Database issue): p. D623-31.

31. Nguyen, B.D., et al., *SOGGY: Solvent-Optimized Double Gradient Spectroscopy for Water Suppression. A Comparison With Some Existing Techniques.* Journal of Magnetic Resonance, 2007. **184**(2): p. 263-74.

32. Chacon, O., et al., *Impairment of D-Alanine Biosynthesis in Mycobacterium smegmatis Determines Decreased Intracellular Survival in Human Macrophages.* Microbiology, 2009. **155**(Pt 5): p. 1440-50.

33. Chacon, O., et al., *Mycobacterium smegmatis D-Alanine Racemase Mutants are Not Dependent on D-Alanine for Growth.* Antimicrobial Agents and Chemotherapy, 2002. **46**(1): p. 47-54.

34. Martinez, d.P.A., et al., *Stereospecificity of Reactions Catalyzed by Bacterial D-Amino Acid Transaminase.* Journal Biological Chemistry, 1989. **264**(30): p. 17784-9.

35. Feng, Z. and R.G. Barletta, *Roles of Mycobacterium smegmatis D-Alanine:D-Alanine Ligase and D-Alanine Racemase in the Mechanisms of Action of and Resistance to the Peptidoglycan Inhibitor D-Cycloserine.* Antimicrobial Agents and Chemotherapy, 2003. **47**(1): p. 283-91.

36. Zygmunt, W.A., *Antagonism of D-Cycloserine Inhibition of Mycobacterial Growth by D-Alanine.* Journal of Bacteriology, 1963. **85**: p. 1217-20.

37. Berney, M., et al., *Regulation of Proline Metabolism in Mycobacteria and Its Role in Carbon Metabolism Under Hypoxia.* Molecular Microbiology, 2012. **84**(4): p. 664-81.

38. Freedberg, W.B., W.S. Kistler, and E.C. Lin, *Lethal Synthesis of Methylglyoxal by Escherichia coli During Unregulated Glycerol Metabolism.* Journal of Bacteriology, 1971. **108**(1): p. 137-44.

39. Kohanski, M.A., D.J. Dwyer, and J.J. Collins, *How Antibiotics Kill Bacteria: From Targets to Networks.* Nature Reviews Microbiology, 2010. **8**(6): p. 423-435.

40. Fonknechten, N., et al., *A Conserved Gene Cluster Rules Anaerobic Oxidative Degradation of L-Ornithine.* Journal of Bacteriology, 2009. **191**(9): p. 3162-3167.

41. Jones, W.M., et al., *The Ubiquitous Cofactor NADH Protects Against Substrate-Induced Inhibition of a Pyridoxal Enzyme.* Protein Science, 1996. **5**(12): p. 2545-2551.

42. Neuhaus, F.C., *D-Cycloserine and O-Carbamyl-D-Serine*, in *Mechanism of Action*, D. Gottlieb and P.D. Shaw, Editors. 1967, Springer Berlin Heidelberg. p. 40-83.

43. Peisach, D., et al., *D-Cycloserine Inactivation of D-Amino Acid Aminotransferase Leads to a Stable Noncovalent Protein Complex with an Aromatic Cycloserine-PLP Derivative.* Journal of the American Chemical Society, 1998. **120**(10): p. 2268-2274.

44. Awasthy, D., et al., *Alanine Racemase Mutants of Mycobacterium tuberculosis Require D-Alanine for Growth and are Defective for Survival in Macrophages and Mice.* Microbiology, 2012. **158**(Pt 2): p. 319-27.

45. Milligan, D.L., et al., *The Alanine Racemase of Mycobacterium smegmatis is Essential for Growth in the Absence of D-Alanine.* Journal of Bacteriology, 2007. **189**(22): p. 8381-8386.

46. Mahapatra, S., et al., *N Glycolylation of the Nucleotide Precursors of Peptidoglycan Biosynthesis of Mycobacterium spp. is Altered by Drug Treatment.* Journal of Bacteriology, 2005. **187**(7): p. 2341-7.

47. de Roubin, M.R., D. Mengin-Lecreulx, and J. van Heijenoort, *Peptidoglycan Biosynthesis in Escherichia coli: Variations in the Metabolism of Alanine and D-Alanyl-D-Alanine.* Journal of General Microbiology, 1992. **138 Pt 8**: p. 1751-7.