

Summer 6-24-2014

Application of Nuclear Magnetic Resonance Based Metabolomics to Study the Central Metabolism of Staphylococci

Bo Zhang

University of Nebraska-Lincoln, bozhangchem@gmail.com

Follow this and additional works at: <http://digitalcommons.unl.edu/chemistrydiss>

 Part of the [Bacterial Infections and Mycoses Commons](#), [Biochemical Phenomena, Metabolism, and Nutrition Commons](#), [Chemistry Commons](#), and the [Investigative Techniques Commons](#)

Zhang, Bo, "Application of Nuclear Magnetic Resonance Based Metabolomics to Study the Central Metabolism of Staphylococci" (2014). *Student Research Projects, Dissertations, and Theses - Chemistry Department*. 46.
<http://digitalcommons.unl.edu/chemistrydiss/46>

This Article is brought to you for free and open access by the Chemistry, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Student Research Projects, Dissertations, and Theses - Chemistry Department by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

APPLICATION OF NUCLEAR MAGNETIC RESONANCE BASED
METABOLOMICS TO STUDY THE CENTRAL METABOLISM OF
STAPHYLOCOCCI

by

Bo Zhang

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

June, 2014

APPLICATION OF NUCLEAR MAGNETIC RESONANCE BASED
METABOLOMICS TO STUDY THE CENTRAL METABOLISM OF
STAPHYLOCOCCI

Bo Zhang, Ph.D.

University of Nebraska, 2014

Advisor: Robert Powers

Metabolomics studies the collection of small molecules (metabolites) involved in enzymatically catalyzed reactions, cell signaling and cellular structure. Perturbations in metabolite concentrations have been used to reflect the activity of corresponding enzymes or proteins. Nuclear magnetic resonance (NMR) spectroscopy is a well-known approach for the structure determination of biological macromolecules. Alternatively, NMR has recently been established as a valuable tool of metabolomics, in which NMR spectral signals correlate small molecules with cellular activities. This has been accomplished through the chemometric analysis of high-throughput one dimensional ^1H spectra (metabolic fingerprinting) and quantitative metabolite identification based on two dimensional ^1H , ^{13}C HSQC spectra (metabolic profiling).

Staphylococcus aureus and *Staphylococcus epidermidis* are the leading pathogens that contribute to a large portion of fatal infections in the USA because of their virulence factors, abilities to survive and thrive on various hosts, and enhanced drug-resistance through biofilm formation. The ability of pathogens to “sense” environmental conditions and to mediate an adaptation of its metabolism to various conditions was studied using NMR-based metabolomics. My research projects included: the correlation between TCA cycle inactivation and biofilm formation related stressors in *S. epidermidis*; investigating

the metabolic details contributing to biofilm formation in *S. epidermidis* under conditions that repress TCA cycle activity; establishing a correlations between iron depletion and oxygen limitation in *S. aureus* metabolism; and evaluation of CcpE as a positive TCA cycle regulator in *S. aureus*.

Ribose phosphate isomerase R (RpiR) is a transcriptional regulatory protein involved in the pentose phosphate pathway. Inactivation of the TCA cycle increases carbon flow into the pentose phosphate pathway where the RpiR protein family may be involved. In *S. aureus*, mutations of intracellular ribose sensing regulators (members of the RpiR family) resulted in changes in the synthesis of virulence factors. The inducer for ribose phosphate isomerase A, B and the mechanism by which RpiRa regulates *rpiA*, *rpiB* gene expression remain to be elucidated. The C-terminal domain of RpiRa was predicted to be a sugar isomerase binding protein domain using homology modeling and it was overexpressed and purified using an *E. coli* pET overexpression system as a first step towards the structural determination of this protein.

ACKNOWLEDGEMENT

If I knew what was ahead of me seven years ago, I probably would have not chosen graduate school. I came into this study program with minimal background of either biology or analytical chemistry. However, when I looked to the past seven years, I realized that what I achieved is amazing. I feel fortunate that I made the right choice and took the first step. But the encouragement from ignorance is not the only thing I need to overcome the real obstacles in every aspect of life in the graduate school. I will not be able to write this thesis without the help from many people shown in my lucky life.

First I would like to thank my advisor, Dr. Robert Powers. He gave me the opportunity to study in his group after the first rotation with the confidence on my interest. He gave me the freedom to develop new ideas, the room to make every possible mistake (and learn from them), and the guidance to struggle in the right direction. These really helped me to grow independently in research and as a human being. His long-term support with his skillful guidance is the key to lead my thesis accomplishment. All the conversations that we have about lab notebooks and safety, plagiarism, English writing as well as “strategies” in the Theme Park of the Disney World will always be cherished in my memory. I would also like to thank my collaborator, Dr. Greg Somerville. He gave me the opportunity to explore microbiology in combination with the development of our NMR-metabolomics methodology. Whenever he said I am wrong about something in biology, it only motivated me to find interesting and exciting aspects of biology. I would like to thank other members of my committee, Dr. David Hage, Dr. Gerard Harbison and

Dr. Hui Li for their advice and support for not only my RUI, OPO and defense, but for my daily research work as well.

It has been a pleasure working with some very talented colleagues in Dr. Powers Lab. I would like to thank Dr. Kelly Mercier, Dr. Matt Shortridge, Dr. Jaime Stark, Dr. Jennifer Copeland, Dr. Steve Halouska, Bradley Worley, Teklab Gebregiworgis, Shulei Lei, Darrell Marshall, and Jonathan Catazaro. The lab is my family in America because of them. I would also consider the personnel in the NMR facility: Dr. Martha Morton, Dr. Joseph Dumais and Sara Basiaga are a part of our group and I owe them a big thank you.

The accomplishment of the thesis is not the work of just one person. I could not have made it this far without all the inspiring conversations and group meeting “punches”. Moreover, I want to thank Dr. Marat R. Sadykov, Dr. Nagender Ledala, Dr. Yefei Zhu and Dr. Rosmarie Gaupp in Dr. Somerville’s lab. They are one of the reasons that I am motivated to be a postdoctoral student. I can talk to them frankly about anything, and they always have an answer or advice for me. One day, I hope to be like them.

Finally I would like to thank my parents and my grandparents. Their health and happiness are the largest support for me thousands of miles away. I also would like to thank the company of my girlfriend, Cong Bi as a strong support in these stressful years. Their understanding and support as I chase my dream is a priceless treasure that I feel most fortunate to have.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION	1
1.1 A brief history of metabolomics.....	1
1.2 Experimental protocols for NMR-based metabolomics.....	4
1.3 Optimization of NMR-based metabolomics methods.....	9
1.4 Metabolism-dependent biofilm formation in <i>S. epidermidis</i> and <i>S. aureus</i>	14
1.5 The RpiRa protein expression system and structural determination.....	17
1.6 Overview of thesis.....	20
1.7 References.....	23
CHAPTER 2: LITERATURE REVIEW	31
2.1 Introduction.....	31
2.2 Biofilm structure.....	34
2.3 Biofilm formation at the molecular level.....	39
2.4 An overview of metabolomics.....	46
2.5 Metabolomics sample preparation.....	55
2.6 NMR experiments for metabolomics.....	58
2.7 Chemometrics and bioinformatics analysis of metabolomics data.....	73
2.8 Analysis of biofilms with NMR-based metabolomics.....	83
2.9 Conclusion and future perspective.....	93

2.10 References.....	96
----------------------	----

CHAPTER 3: REVISITING PROTOCOLS FOR THE NMR ANALYSIS OF BACTERIAL CENTRAL METABOLOMES.....120

3.1 Introduction.....	120
-----------------------	-----

3.2 Experimental design.....	123
------------------------------	-----

3.2.1 Identify the appropriate biological system for a metabolomics study.....	125
--	-----

3.2.2 Minimize unintended bias and biologically irrelevant variations.....	127
--	-----

3.3 Sample preparation.....	129
-----------------------------	-----

3.3.1 Bacterial cultivation.....	129
----------------------------------	-----

3.3.2 Quenching, washing and harvesting the bacterial cells.....	134
--	-----

3.3.3 Cell lysing and metabolite extraction.....	138
--	-----

3.3.4 Isotopic labeling of metabolites.....	139
---	-----

3.3.5 Storage of metabolite samples.....	141
--	-----

3.4 NMR spectroscopy.....	142
---------------------------	-----

3.4.1 One-dimensional ^1H NMR methodology.....	142
---	-----

3.4.2 Two-dimensional ^1H , ^{13}C HSQC NMR methodology.....	147
--	-----

3.5 Data analysis.....	148
------------------------	-----

3.5.1 Preprocessing of 1D ^1H NMR data.....	148
--	-----

3.5.2 Multivariate statistical analysis of 1D ^1H NMR data.....	154
--	-----

3.5.3 Metabolite identification.....	156
--------------------------------------	-----

3.5.4 Metabolomics network map.....	165
-------------------------------------	-----

3.6 Conclusion.....	168
---------------------	-----

3.7 References.....	169
---------------------	-----

CHAPTER 4: AN INEXPENSIVE HIGH-THROUGHPUT NMR TUBE

CLEANING APPARATUS.....	179
4.1 Introduction.....	179
4.2 Methods and materials.....	181
4.3 Conclusion.....	186
4.4 References.....	186

CHAPTER 5: USING NMR METABOLOMICS TO INVESTIGATE

TRICARBOXYLIC ACID CYCLE DEPENDENT SIGNAL TRANSDUCTION IN <i>STAPHYLOCOCCUS EPIDERMIDIS</i>.....	188
5.1 Introduction.....	188
5.2 Methods and materials.....	192
5.2.1 Bacterial growth and NMR sample preparation.....	192
5.2.2 NMR data collection.....	194
5.2.3 NMR data analysis.....	195
5.3 Results.....	197
5.3.1 NMR metabolomics and principal component analysis.....	197
5.3.2 Harvesting of <i>S. epidermidis</i> cultures.....	198
5.3.3 Impact of environmental stress conditions on the <i>S. epidermidis</i> metabolome.....	207

5.3.4 Detailed analysis of changes to the <i>S. epidermidis</i> metabolome caused by environmental stress.....	217
5.3.5 Metabolic rearrangements during TCA cycle stress.....	226
5.4 Discussion.....	228
5.5 References.....	228
5.6 Appendix A: Effect of autoinducer-2 (AI-2) on the <i>S. epidermidis</i> metabolome.....	236

CHAPTER 6: NMR ANALYSIS OF A STRESS RESPONSE METABOLIC SIGNALING NETWORK.....	239
6.1 Introduction.....	239
6.2 Methods and materials.....	241
6.2.1 Bacterial strains, media, and growth conditions.....	241
6.2.2 Aconitase activity assay.....	242
6.2.3 Northern blot analysis.....	242
6.2.4 PIA immunoblot assay.....	244
6.2.5 NMR sample preparation.....	244
6.2.6 NMR analysis.....	245
6.2.7 Metabolomic dendrogram.....	247
6.3 Results.....	248
6.3.1 Disparate environmental stresses create a metabolic block in the TCA cycle.....	248
6.3.2 Environmental stimuli elicit TCA cycle-dependent metabolic changes.....	249
6.3.3 Metabolomic changes are largely independent of the σ^B -mediated general stress response.....	258

6.3.4 Metabolomic changes precede genetic changes.....	259
6.3.5 CcpA responds to TCA cycle-associated metabolomic changes.....	261
6.3.6 TCA cycle stress decreases RNAIII transcription.....	263
6.4 Discussion.....	264
6.5 Conclusion.....	267
6.6 References.....	267
CHAPTER 7: INFLUENCE OF IRON AND AERATION ON <i>STAPHYLOCOCCUS AUREUS</i> GROWTH, METABOLISM, AND TRANSCRIPTION.....	274
7.1 Introduction.....	274
7.2 Methods and materials.....	277
7.2.1 Bacterial strains, media, and cultivation conditions.....	277
7.2.2 Sample preparation for NMR metabolomic analysis.....	278
7.2.3 Data collection.....	279
7.2.4 Data analysis.....	279
7.2.5 Data interpretation.....	280
7.2.6 ICP-MS analysis of DTSB.....	280
7.2.7 Aconitase activity assay.....	281
7.2.8 Real-time RT-PCR.....	282
7.2.9 Acetate, glucose and lactate level determination of the media supernatant...	283
7.3 Results.....	283
7.3.1 Growth of <i>S. aureus</i> under iron- and oxygen-limitation.....	283
7.3.2 Iron- and oxygen-limitation create metabolic blocks.....	287

7.3.3 Iron and/or oxygen limitation alters the <i>S. aureus</i> metabolome.....	289
7.3.4 Iron-regulated gene transcription varies with oxygen availability.....	296
7.4 Discussion.....	297
7.5 References.....	307

CHAPTER 8: CATABOLITE CONTROL PROTEIN (CCPE) IS A LYSR-TYPE TRANSCRIPTIONAL REGULATOR OF TRICARBOXYLIC ACID CYCLE ACTIVITY IN <i>STAPHYLOCOCCUS AUREUS</i>.....	313
8.1 Introduction.....	313
8.2 Methods and materials.....	314
8.2.1 Bacterial strains and culture conditions.....	314
8.2.2 Construction of a <i>S. aureus</i> Δ <i>ccpE</i> mutant.....	317
8.2.3 Construction of the <i>S. aureus</i> <i>ccpE</i> cis-complementation strain TH01c.....	319
8.2.4 Construction of the <i>B. subtilis</i> <i>ccpC</i> promoter - <i>B. subtilis</i> <i>ccpC</i> and <i>B. subtilis</i> <i>ccpC</i> promoter - <i>ccpE</i> trans-complementation plasmids.....	319
8.2.5 Aconitase activity assay.....	320
8.2.6 Citrate synthase activity assay.....	320
8.2.7 Determination of citrate by GC/MS.....	321
8.2.8 Transcriptional analyses.....	322
8.2.9 Determination of transcriptional start sites of NMMN_0640 and <i>citB</i>	323
8.2.10 Antibody production and immunoblotting.....	323
8.2.11 Electrophoretic mobility shift assays.....	324
8.2.12 NMR sample preparation.....	325
8.2.13 NMR data collection, analysis and interpretation.....	325

8.2.14 Statistical analyses.....	325
8.3 Results.....	325
8.3.1 Identification of potential carbon metabolism-affecting factors in <i>S. aureus</i> strain Newman.....	326
8.3.2 Transcriptional organization of the <i>ccpE</i> locus in <i>S. aureus</i> strain Newman.....	326
8.3.3 Inactivation of <i>ccpE</i> in <i>S. aureus</i> strain Newman.....	332
8.3.4 CcpE affects the in vitro growth yield of <i>S. aureus</i>	333
8.3.5 Deletion of <i>ccpE</i> alters the <i>S. aureus</i> metabolome.....	334
8.3.6 Inactivation of <i>ccpE</i> affects transcription of TCA cycle genes and its activity.....	337
8.3.7 CcpE binds to the <i>citB</i> promoter.....	340
8.3.8 CcpE is not a functional homolog of CcpC.....	340
8.4 Discussion.....	343
8.5 References.....	345
CHAPTER 9: OVEREXPRESSION AND PURIFICATION OF ISOTOPICALLY LABELED C-TERMINAL DOMAIN OF RPIRA FOR NMR STUDIES.....	349
9.1 Introduction.....	349
9.2 Methods and materials.....	356
9.2.1 Materials.....	356
9.2.2 Plasmid construction, <i>E coli</i> . strains and glycerol starter.....	357
9.2.3 Protein expression in M9 medium and high cell density medium.....	358
9.2.4 Protein extraction and purification.....	359

9.2.5 Imidazole removal, protein concentration and NMR sample preparation.....	360
9.2.6 Analytical apparatus and data analysis.....	360
9.3 Results.....	362
9.3.1 Determination of CTD amino acid region.....	362
9.3.2 Protein expression.....	365
9.3.3 Protein extraction.....	368
9.3.4 Protein purification.....	369
9.3.5 Protein buffer conditions.....	371
9.3.6 Isotopic labeling of RpiRa CTD.....	374
9.4 Discussion.....	378
9.5 References.....	379
Appendix B: Medium and buffer recipes.....	383
CHAPTER 10: SUMMARY AND FUTURE DIRECTIONS.....	387
10.1 Summary.....	387
10.2 The influence and future directions of metabolomics in solution state NMR.....	391
10.3 References.....	393

PUBLICATION LIST

The thesis includes chapters adapted from published manuscripts (Chapter 2-8) in peer reviewed journals as listed below:

Chapter 2:

B. Zhang, R. Powers*, (2012) "Analysis of Bacterial Biofilms Using NMR-based Metabolomics" *Future Medicinal Chemistry* 4 (10):1273-1306

Chapter 3:

S. Halouska[†], B. Zhang[†], R. Gaupp, S. Lei, E. Snell, R. J. Fenton, R. G. Barletta, G. Somerville, R. Powers*, (2013) "Revisiting Protocols for the NMR Analysis of Bacterial Metabolomes" ([†] equal contribution) *Journal of Integrated Omics* 3(2):120-137

Chapter 4:

B. Zhang, J. Hodgson, W. Hancock, R. Powers*, (2011) "An Inexpensive High-Throughput NMR Tube Cleaning Apparatus" *Analytical Biochemistry* 416 (2): 234-236

Chapter 5:

M. R. Sadykov, B. Zhang, S. Halouska, J. L. Nelson, L. W. Kreimer, Y. Zhu, R. Powers, G. A. Somerville*, (2010) "Using NMR metabolomics to investigate tricarboxylic acid cycle dependent signal transduction in *Staphylococcus epidermidis*." *Journal of Biological Chemistry* 285(47):36616-36624

Chapter 6:

B. Zhang, S. Halouska, C. E. Schiaffo, M. R. Sadykov, G. A. Somerville, R. Powers*, (2011) "NMR Analysis of a Stress Response Metabolic Signaling Network" *Journal of Proteome Research* 10 (8): 3743-3754

Chapter 7:

N. Ledala[†], B. Zhang[†], G. A. Somerville, R. Powers*, (2014) "The impact of iron and oxygen limitation on the metabolome of *Staphylococcus aureus*" ([†] equal contribution) *Journal of Bacteriology* 196(12):2178-2189

Chapter 8:

T. Hartmann, B. Zhang, G. Baronian, B. Schulthess, S. Grubmuller, E. Lacknermeier, R. Gaupp, R. Bertram, R. Powers, W. Eisenreich, M. Hermann, V. Molle, G. A. Somerville, M. Bischof*, (2013) "Catabolite Control Protein E (CcpE) is a LysR-type Regulator that Controls TCA Cycle Activity in *Staphylococcus aureus*" *Journal of Biological Chemistry* 288 (50):36116-36128

“If I were doing a PhD, I’d be doing it in Metabolomics.”

- James Watson (2013)

CHAPTER 1

INTRODUCTION

1.1 A brief history of metabolomics

Metabolomics is a new member of the “omics” community and focuses on the analysis of a collection of small molecules in a biological system.¹ In the past decade, it has gained great attention in fundamental science and has been applied in clinical practice and in the pharmaceutical industry.² The number of papers with the keyword “metabolomics” has grown exponentially in the past decade (Figure 1). As a methodology, metabolomics has made a fundamental impact on analytical techniques such as NMR, mass spectroscopy and chromatography.^{3,4}

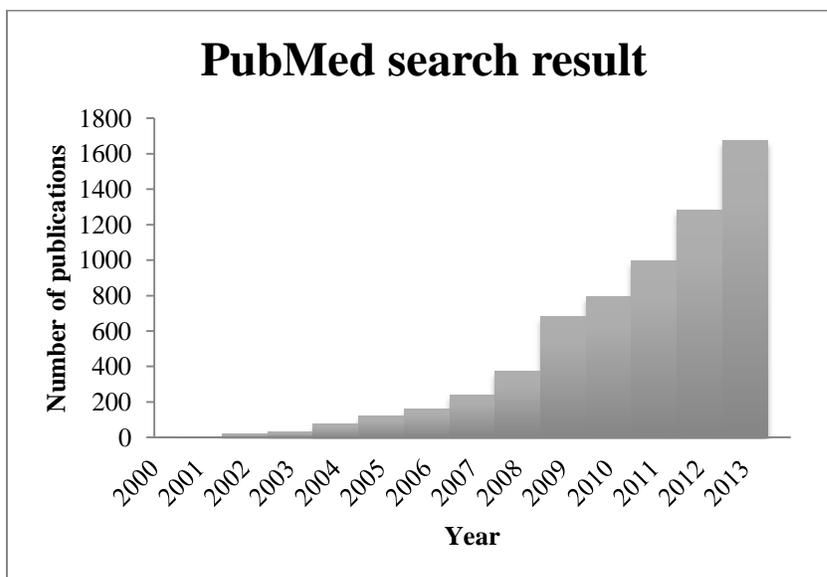


Figure 1.1 PubMed search result of the number of publications with “metabolomics” as a keyword.

Metabolites can be defined as all the small molecules with a molecular weight less than 1 kDa subject to enzymatic transformation in all metabolic pathways.⁵ They can be categorized into intracellular/intercellular metabolites based on the location of their inception. The roles that metabolites play in a biological system vary greatly. First, they are small molecules that can be used as building blocks by numerous metabolic pathways to synthesize biological macromolecules. Second, they can be used by organism-dependent respiration processes to provide energy in order to maintain regular cellular functions. Furthermore, because of their high turnover rate, they are ideal molecules for cells to produce and consume so as to counterbalance the environmental stimuli and keep a homeostasis within the cells.⁶ A growing amount of evidence has shown that the mediation of gene transcription and protein expression by metabolites is not rare, but more likely to be a universal phenomenon.^{7,8} As metabolites participate in many cellular activities, they can be used to evaluate the activity of proteins. Similarly, all of the metabolites within a cell, a tissue, an organ, or even an organism can be used to represent the biological state of the system.

The complete collection of metabolites is termed as the metabolome. The concept of studying a large set of metabolites simultaneously under a certain set of conditions is called “metabonomics” or “metabolomics”, representing two viewpoints of “how” to interpret the metabolome.⁹ Professor Jeremy Nicholson, a pioneer of the field, suggested the term “metabonomics” to be used during measurement of the multivariate metabolic responses to various environmental and toxic stimuli.⁹ The metabolome can be therefore used for biomarker identification¹⁰ and drug toxicity evaluation¹¹ among other

examples.^{12,13} NMR is mainly used for such applications in “metabonomics” and is ideal for the detection of amino acids, organic acids, amines, carbohydrates and lipids. Another camp uses the term “metabolomics”, which was first coined by Professor Oliver Fiehn.¹⁴ The information obtained using “metabolomics” places a greater emphasis on metabolic profiling at a cellular or organ level and is primarily concerned with normal endogenous metabolism. Genomics and proteomics data are routinely combined and used as a tool to study functional genomics.^{15,16} Mass spectrometry (MS) has been heavily employed in this camp to detect the metabolome.¹⁷ MS has to be coupled with gas chromatography, liquid chromatography or capillary electrophoresis because of the low molecular-weight dispersion of the metabolome and because biologically irrelevant information introduced from ion suppression, matrix effect, and analyte fragmentation can otherwise prevent the results from representing the true metabolome *in vivo*.^{18,19} Liquid Chromatography (LC)-MS can be used to identify amino acids, amines, fatty acids, nucleosides, lipids, carbohydrates. Gas Chromatography (GC)-MS can be used for organic acids, aldehydes, ketones, volatiles, fatty acids, amino acids, and steroids. In practice, the difference between “metabolomics” and “metabonomics” has become too obscure, so the terms are used interchangeably in the literature, and as a matter of personal preference.

Different instruments have their own strengths and weaknesses and, correspondingly, different coverage of the metabolome, so the correct choice of instrumentation is critical to successfully address the specific metabolic question. The selection of proper instrumentation is dependent on the availability of these platforms, as well as the ability to combine different instruments to obtain a comprehensive metabolic

profile. NMR is known for its simple sample preparation, high reproducibility and ease of quantification so it is popular in detecting primary metabolites and metabolic fingerprints. MS is well-established for detecting secondary or low abundant metabolites with its high sensitivity in conjunction with a chromatography system. Currently, both instruments have been widely used as standalone systems, or in combination, to analyze tissues,²⁰ biofluids,²¹ mammalian cell culture,²² and plants.²³ As an illustration, a recently completed study identified a total of 445 metabolites in human urine samples using a multi-platform analysis: 209 metabolites by NMR, 179 metabolites by GC-MS, 127 metabolites by DFI/LC-MS/MS, 40 metabolites by ICP-MS and 10 by HPLC.²⁴ Thus, platforms that combine multiple analytical techniques are likely to become a standard approach in metabolomics. About 50 metabolites were regarded as the “cellular metabolome” for *E. coli* in 1972. In 2013, a metabolomics database contains more than 2600 metabolites for the basic K-12 strain of *E.coli*.²⁵ Thus, advancements in instrumentation and the development of new methodologies are currently benefiting metabolomics and are key factors to continued progress in the field.

1.2 Experimental protocols for NMR-based metabolomics

Metabolome samples can be generated from bacterial strains, human cell lines, animal tissues, and various biofluids (urine, serum, *etc.*). The preparation protocols for these different sources are rather diverse due to sample variability in pH, salt concentration, and protein content, among numerous other factors. An efficient metabolome sampling procedure also needs to consider the properties of the metabolites

of interest: size, structure, solubility, concentration and stability. In addition, the response of the metabolome to environmental or experimental changes can be extremely rapid. Therefore, accurately analyzing the metabolome to obtain biologically relevant information is technically very difficult and requires extensive methodology development. In our efforts to establish a successful metabolomics sample preparation workflow, three standards were considered.

The first standard considered was efficiency because of the relatively low sensitivity of NMR. It is particularly challenging to completely extract the entire metabolome from a cell, tissue, organ or organism. So a series of “trial and error” experiments are required to optimize cell lysis, extraction buffer, and instrumental parameters. A second critical criterion of metabolomics is the need to ensure that the sample represents the “true” state of the cell or system being studied. The key is to process all of the samples as quickly as possible while also stopping or slowing down all cellular activities by quenching the sample. Achieving these goals makes it likely that an accurate “snapshot” of the metabolome has been obtained. Consequently, quenching cellular activities is a critical step of properly preparing metabolomics samples, where evaluating different extraction and quenching procedures are discussed in Chapter 3. Consistency was the third standard considered. The primary goal is to avoid the introduction of biologically irrelevant variations in the metabolome by the investigator or by any step of the protocol. While the introduction of irrelevant variations in metabolomics samples is unavoidable, the goal is to optimize experimental protocols in order to drastically reduce their occurrence and negative impact. An advantage of NMR

is its easy sample preparation that does not require any chromatography or sample separation step, which effectively removes variations attributed to sample handling and preparation. Similarly, it is also critical to include proper experimental design in regards to data collection and data process to reduce unwanted variation or bias and to ensure that the variations detected are limited to the factors being modified (eg. untreated vs. treatment). Besides these general standards, each research projects required further optimization of sample preparation, metabolome extraction, data collection and data analysis procedures. Improving the reproducibility and the quality of NMR-based metabolomics data were important components of each project. Thus, additional advancements in our metabolomics protocols and experimental details that led to improved metabolomics datasets are described throughout the thesis.

The identification and quantification of metabolites is an important component of metabolomics, where there are two general strategies – targeted and untargeted metabolomics. In targeted metabolomics, the goal is to quantify changes to a specific set of metabolites that have been previously associated with a particular biological process or function.²⁶ In untargeted metabolomics (also referred to as metabolic profiling, metabolic fingerprinting (intracellular metabolites) or footprinting (intercellular metabolites), it is not necessarily required to identify the metabolites. Instead, the goal is to identify global similarities or differences in the metabolomes from two or more groups or classes (wild-type vs. mutant, untreated vs. drug treated, *etc.*) In untargeted metabolomics, the NMR or MS spectra are often subjected to chemometric analysis, such as Principal Component Analysis (PCA) and Orthogonal Projection Latent Structure for Discriminant Analysis

(OPLS-DA).^{27,28} The resulting scores plot (Figure 3) identifies the relative class differentiation based on the clustering pattern. Spectra that cluster together in the scores plot have a similar metabolome. Conversely, the further two clusters are separated in the scores plot indicates a relative difference in the metabolomes. Figure 3 illustrates theoretical PCA scores plots demonstrating how cluster analysis can be used to determine *in vivo* efficacy and selectivity of a drug candidate.²⁹

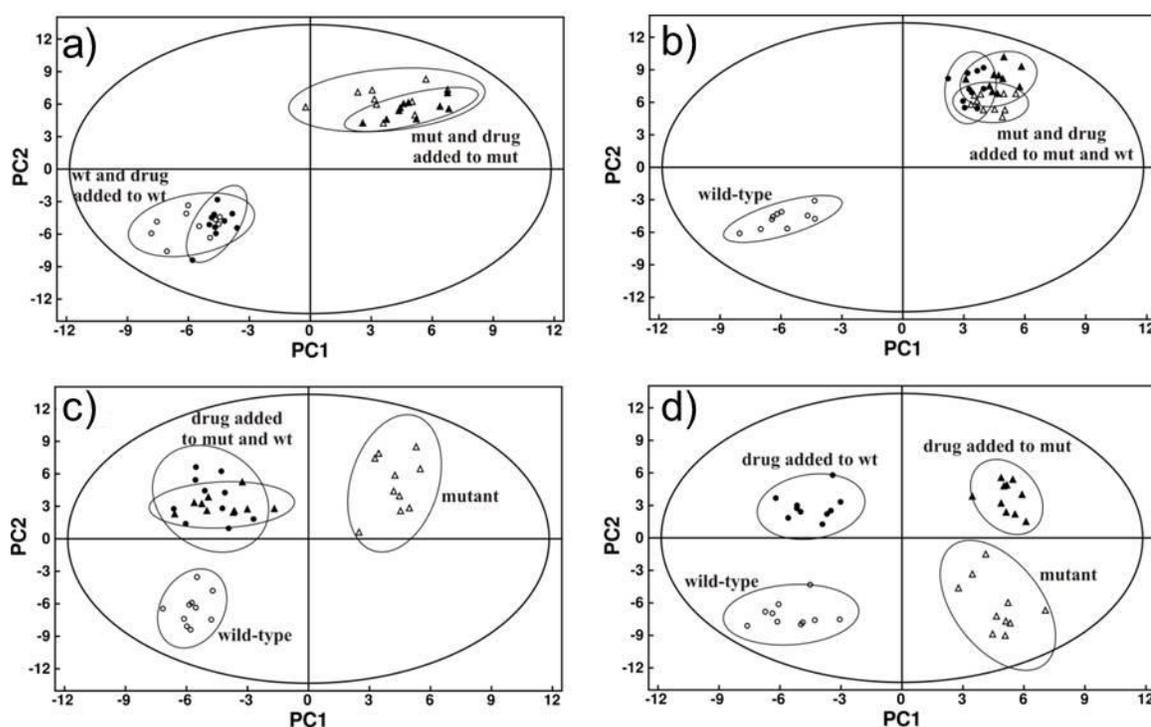


Figure 1.2 Illustration of hypothetical PCA scores plot for the following scenarios a) inactive compound, b) active and selective inhibitor, c) active, nonselective inhibition of target and secondary protein, and d) active, nonselective preferential inhibition of secondary protein. Reprinted with permission from reference.²⁹

In the current application, the metabolome is perturbed by chemicals (i.e. drugs), by biology (i.e. mutations), by phenotype (i.e. disease), or by the environment (i.e. nutrients). The corresponding changes in the metabolome are used to investigate the system's response to the perturbation and identify the underlying molecular mechanisms. For example, pathogenicity related stressors can reveal which normal metabolic processes are connected with virulence biosynthesis. Alternatively, genes with unknown function can be annotated by comparing changes in the metabolome between wild-type and mutant cell lines. Specifically, I applied “differential metabolomics” or “comparable metabolomics” to investigate biological process associated with the survivability and adaptability of staphylococcus bacteria. The quantitative and qualitative analysis of metabolites from intracellular extracts can be extremely valuable when investigating the mechanism of drugs, diseases or environmental stimuli.³⁰⁻³²

NMR based metabolomics can also be used to study the gene functions besides the global monitoring of metabolic perturbations under a stress. Enzymes are the driving force of metabolic reactions and they make up to 25% of genes in microbes and 10% in humans.^{33,34} An *in silico* genome-scale reconstruction of the metabolic network of *Staphylococcus aureus* listed 774 reactions, representing 23% of its open reading frame.³⁵ Unfortunately, our understanding of the metabolism on the system level is far from complete because a large portion of genes encode proteins of unknown function (also known as hypothetical proteins).³⁶⁻³⁹ There are two ways of studying gene functions in terms of metabolomics data: forward genetics and reverse genetics.¹⁴ The forward genetics approach has a defined biological state but unknown perturbation in genes,

proteins and metabolites. The goal is to identify which genes, proteins or metabolites are affected. Reversed genetics refers to the process where a target gene or protein is up- or down-regulated, but its influence has not been explored.⁴⁰ Nevertheless, there is no consensus on how metabolomics data can be used to interpret how a biological systems response to a perturbation. The challenge arises from the fact that individual metabolites can be involved in many known or unknown metabolic pathways. Based on a single concentration change, it is difficult to accurately distribution this change through multiple pathways.

Another system-level approach to understand metabolomics data is computational modeling.⁴¹ Flux analysis has been performed by using metabolic control analysis (MCA), flux balance analysis (FBA) and other techniques.⁴²⁻⁴⁵ To date, all flux analysis studies have concluded that it is not a trivial task to understand the “function” of individual metabolite. In other words, there is no definite destination for a metabolite given a complex array of highly-interrelated metabolic pathways. In effect, any solution is strictly case-dependent. This is also a challenge in biomarker discovery.⁴⁶ Thus, taking full-advantage of the power of metabolomics is highly dependent on establishing a restricted and determinant state, where genomics, transcriptomics, and proteomics data can provide a biological context to the metabolomics data. In effect, the “omics” data complement each other.

1.3 Optimization of NMR-based metabolomics methods

Statistical methods, especially multivariate regression methods that include PCA and OPLS, are routinely used in the analysis of one dimensional (1D) NMR spectra. Therefore, a large portion of my research efforts were dedicated to the optimization of PCA analysis of NMR datasets. In a PCA scores plot generated from NMR metabolomics data, the analysis mainly lies in the determination of the relative distances between multiple “clusters”. The presence of different class clusters is due to the fact that within-group variability is smaller than the between-group variability. Multiple sample replicates are required to statistically define a class cluster, which are dependent on the reproducibility of the replicates. In practice, within-group variability is defined by a combination of experimental conditions and the inherent variability of the biological sample. Thus, a larger data set is always desirable because of the statistical power it possesses in defining statistically relevant class separation.⁴⁷ To address the need to determine the statistical significance of class differentiation in a scores plot, a tree-diagram based on a matrix of center-averaged distances between each cluster was generated. The significance of each node is based on traditional boot-strapping approaches. The tree-diagram provides a quantitative alternative to the direct inspection of a PCA scores plot.⁴⁸

Heteroscedasticity is a commonly neglected property of chemometric data and it has a negative impact on the multivariate classification. Simply, the variance of a variable increases as the value of a correlated variable increases. This leads to biased standard errors and biased inferences. Biological variability is one source of heteroscedasticity. For example, more metabolic pathways are active during exponential growth when

reproduction is taking place compared to post-exponential growth when metabolism is focused on survival due to limited carbon sources. Therefore, the dynamics and composition of the metabolome is completely different between the two phases. As a result, metabolomes extracted from cells during the post-exponential phase resulted in tighter clusters in a scores plot. Currently, there are few applications that explore such information and Dr. Daniel Raftery is a leading researcher in this area. Ratio Analysis of NMR Spectroscopy (RANSY) and Ratio Analysis of Mass Spectrometry (RAMSY) have been developed based on the assumption that individual peaks can be correlated by their relative standard deviation.^{49,50} For each individual metabolite in a sample set, there can only be one relative standard deviation in the metabolites concentration and, therefore, a correlation between relative standard deviations can be used to assign all the NMR peaks originating from the same metabolite. This assumes a small likelihood that peaks originating from different metabolites having the same relative standard deviation. This method is complementary to their statistical total correlation spectroscopy (STOCSY) method. STOCSY is based on a correlation of peak intensity that assumes coupling patterns and peak overlap does not mask the correlated peaks.⁵¹ We have also taken advantage of heteroscedasticity in our methodology development, albeit from a different direction. The random, low intensity peaks in a 1D ^1H NMR spectrum are considered noise. It was previously shown that the inclusion of noise in a multivariate analysis can negatively impact clustering in a scores plot and lead to erroneous interpretations.⁵² But, the complete removal of noise regions, while leaving real signals intact, is particularly challenging for complex NMR spectra of biological samples. To resolve this issue, a

novel noise removal method was developed based on the relative standard deviation of noise instead of the traditional cutoff based on arbitrary minimum signal intensity. A region absent of peaks (spectral edge) is used to define the “noise” parameters, which is then used to remove all noise from the remainder of the spectrum. The removal of noise ensures that our PCA analysis does not unintentionally mask any biologically relevant trends.

Two-dimension (2D) ^1H , ^{13}C Heteronuclear Single Quantum Coherence (HSQC) correlates protons attached directly to a carbon, which is a very basic component of organic structures. Thus, the 2D ^1H , ^{13}C HSQC experiment is a valuable approach to improve the accuracy in metabolite identification. Although the analysis of 2D NMR datasets takes considerable more time and effort, 2D NMR experiments are increasingly being used by the metabolomics community because of several distinct advantages. The two chemical shifts (^1H and ^{13}C) from a C-H pair must both match the chemical shift values in a NMR metabolomics database to make a metabolite assignment. Also, peak overlap and chemical shift ambiguity is significantly reduced by dispersing the chemical shift information into two-dimensions. The 2D ^1H , ^{13}C HSQC experiment has been recently improved and optimized for metabolomics to directly measure concentrations (Figure 3).^{53,54} To enhance the detection of metabolites,⁵⁵ the culture media is supplemented with a ^{13}C -labeled compound to boost the sensitivity of the 2D ^1H , ^{13}C HSQC experiment due to the low natural abundance (1.1%) of ^{13}C .^{56,57} More importantly, only metabolites that are derived from the ^{13}C -labeled compound are observable in the 2D ^1H , ^{13}C HSQC spectrum. It allows for monitoring carbon flow from the ^{13}C -labeled

compound throughout the metabolome. This also simplifies the analysis by restricting the metabolome to a select sub-set of the metabolites. The judicious choice of the ^{13}C -labeled compound will only emphasize the region of the metabolome that is particularly relevant to the biological system under investigation. The application of 2D ^1H , ^{13}C HSQC spectroscopy has been largely incorporated into our NMR-based metabolomics, which includes a method to automate peak alignment, peak-picking, absolute concentration determination, and multiple ways of representing metabolome changes.

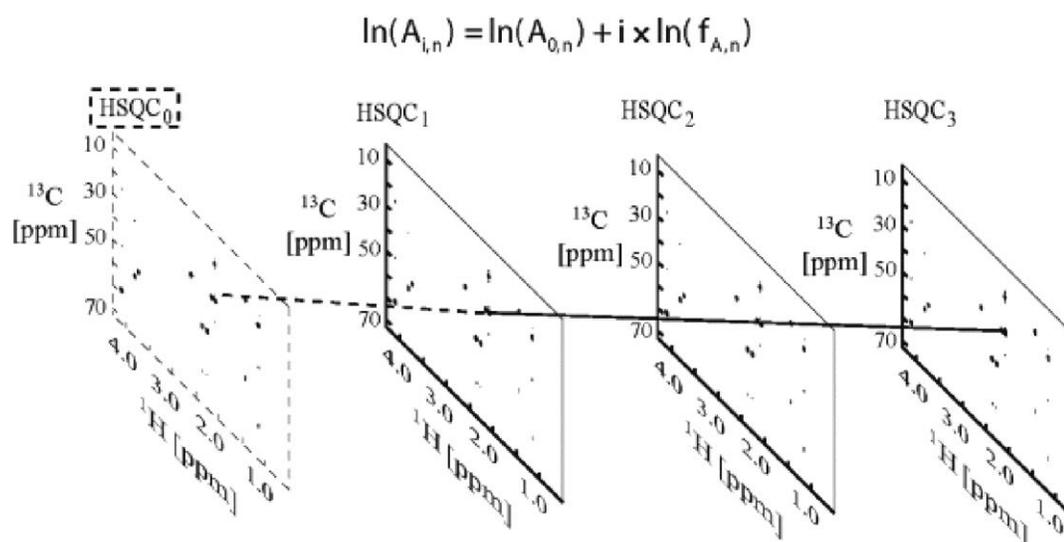


Figure 1.3 The principle of Time-0 HSQC experiment. A series of HSQC spectra acquired with incremented repetition times (the time between the end of the first ^1H excitation pulse to the beginning of data acquisition) can be extrapolated back to zero time to yield a time-zero 2D ^1H , ^{13}C HSQC spectrum (HSQC₀) in which signal intensities are proportional to concentrations of individual metabolites. Reprinted with permission from reference^{53,54}

1.4 Metabolism-dependent biofilm formation in *S. epidermidis* and *S. aureus*

Biofilms are a ubiquitous microbial phenomenon that has been observed for bacteria, fungi, algae, yeasts, protozoa, and other organisms.⁵⁸⁻⁶⁰ Biofilms were first observed in 1923 during the fouling of ships' hulls by marine microorganisms.⁶¹ In 1933, Henrici wrote "it is quite evident that for the most part water bacteria are not free floating organisms, but grow upon submerged surfaces".⁶² In 1940, Heukelekian and Heller added "development takes place either as bacterial slime or colonial growth attached to

surfaces".⁶³ It is worth noting that the extensive interest in biofilms are not merely a result of scientific curiosity, but are also derived from practical concerns related to a broad spectrum of areas such as medical science,⁶⁴ material engineering,⁶⁵ civil engineering,⁶⁶ and others.⁶⁷ In particular, biofilm-related bacterial infections are a serious health concern. A 2000 US government report identified infectious diseases as a leading cause of death worldwide and the third leading cause of death in the United States.⁶⁸ It has been estimated that 60%-80% of human microbial infections are caused by bacteria growing as a biofilm.⁶⁹ Certain pathogenic biofilms are particularly concerning because of the added issue of drug-resistance.⁷⁰ As methicillin-resistant *Staphylococcus aureus* (MRSA) and other resistant pathogens capable of biofilm formation continue to emerge and propagate, understanding and circumventing biofilm resistance to antibiotics is a paramount necessity.⁷¹

Bacteria can respond to a universal signal of nutrient exhaustion by 1) swimming to new locations; 2) synthesizing enzymes that degrade macromolecules and use the products; 3) secreting antibiotics that kill competitor bacteria in the local environment. A biofilm is likely to form if these strategies are not successful in promoting normal culture growth in a harsh environment.⁷² The biofilm phenotype is not exclusively caused by this mechanism.⁷³ In a biofilm, the metabolism of the bacteria is dramatically lowered so that a minimal amount of nutrients are required for survival. Thus, a biofilm may provide further protection against antibiotics that typically target cell growth-related processes. Also, the biofilm matrix may provide a further layer of antibiotics resistance by acting as a physical barrier and hindering drug delivery to the bacteria.⁷⁴ Intercellular

communications within a biofilm and between planktonic cells is important to our understanding of bacterial behavior.⁷⁵ Quorum sensing,⁷⁶ programmed cell death or population-wide drug resistance may be regulated by or are established examples of intercellular communications.^{65,66} These observations raise some questions about the validity of research into bacterial behavior that is based solely on highly homogenous cultures. At a minimum, it suggests the need for studies that involve both planktonic and biofilm forms in order to obtain a complete picture of bacterial cellular biology. In this sense, the study of bacterial biofilms is still in its infancy.

As with most pathogens, the survival of staphylococcal bacteria is largely dependent on host-pathogen interactions or on the environment. The host environment provides the necessary nutrients and energy for the bacteria to grow and reproduce. These cellular activities are accomplished through a series of metabolic pathways. In staphylococci, the central metabolic pathways that include glycolysis, pentose phosphate pathway and the tricarboxylic acid cycle (TCA cycle) can produce all the intermediate building blocks for macromolecular production.⁷⁷ To be more specific, the TCA cycle is a major source of the metabolites and energy for the bacteria. Biofilm formation, as a survival strategy, is also largely mediated by environmental stimuli and nutrient availability (Figure 4). Thus, bacteria respond to a multitude of stressors to induce a biofilm. Thus, based on its major role in central metabolism, the TCA could translate environmental stressors into the metabolic signals and subsequently control the metabolism and pathological activities of the cells.⁷ An accumulating amount of evidence has demonstrated that cell sensing and regulation, detoxification, repair and protection,

Studying metabolism is challenging because it is one of the most flexible and complex systems of a living organisms in terms of regulation. Metabolism connects all of the components (*e.g.*, proteins, genes) of a cell with the external environment. As the number of metabolites identified continues to increase, an obvious question arises - what are the biological functions of these metabolites? Determining the structure of a protein may be important to addressing these questions. The availability of a protein structure may assist in our understanding of the protein's biological function, which may include *interactions with small molecules like metabolites*. Conversely, identifying metabolites that interact with a protein may establish the biological importance of a specific metabolite.

TCA cycle stress alters the intracellular ribose concentration in *S. epidermidis* and also alters the temporal expression of virulence factors in *S. epidermidis* and *S. aureus*. One metabolic change associated with TCA cycle repression is an increased concentration of intracellular ribose. It indicates the presence of a pentose phosphate pathway (PPP)-responsive regulator that can partially mediate TCA cycle-dependent regulatory effects.⁷⁷ In biofilm formation, the replication and production of nucleotides requires overproduction of ribose. As a result, a PPP-responsive regulator may mediate some of the TCA cycle-dependent regulatory effects.

Ribose Phosphate Isomerase (Rpi) is a family of protein hypothetically mediating the regulation of PPP. A search of the *S. aureus* strain Mu50 genome returned three open reading frames (ORFs) with significant amino acid homology to RpiR (21 to 23% amino acid identity and 45 to 46% amino acid similarity): SAV0317, SAV0193, and SAV2315.

For simplicity, these homologues were designated RpiRa (SAV0317), RpiRb (SAV0193), and RpiRc (SAV2315).⁸⁰ Two of these three homologues of Rpi (RpiRb and RpiRc) have been found to positively influence the transcription of the PPP genes *rpiA* and *zwf*, while the third homologue (RpiRa) is slightly antagonistic.^{80,81} In addition, inactivation of RpiRc altered the temporal transcription of RNAIII, the effector molecule of the agr quorum-sensing system.⁸⁰ Thus, these data reveal the close linkage of central metabolism and virulence determinant synthesis in terms of protein and metabolism regulation.

These observations led us to hypothesize that an RpiR homologue may link the PPP to virulence factor regulation in staphylococci. The structure determination is a direct way to understand protein functions in detail. Currently there is only one X-ray crystal structure of RpiR C-terminal domain (CTD) from *Sphaerobacter thermophilus* deposited in the Protein Data Bank (PDB) (unpublished data, PDB ID: 3SHO). It has a sequence identity of 22% and similarity of 44% with RpiRa from *S. aureus*. A homology model for the sugar isomerase-like CTD of *S. aureus* RpiRa was built up and examined based on the *Vibrio vulnificus* NanR protein (PDB ID: 4IVN) using sequence alignment tools of FASTA⁸² and BLAST[®]; domain prediction software packages of DomPred⁸³ and InterProScan⁸⁴; secondary structure prediction tools of PredictProtein⁸⁵ and PSIPRED⁸⁶. The CTD of RpiRa, a hypothetical sugar isomerase domain, was proposed to be a ligand binding domain that can sense the perturbation of metabolites. Therefore, it can be used as a metabolite regulatory mechanism in pathogenesis development. The purification of various constructs of RpiRa CTD and analysis of NMR sample conditions to identify a suitable sample for NMR structure determination.

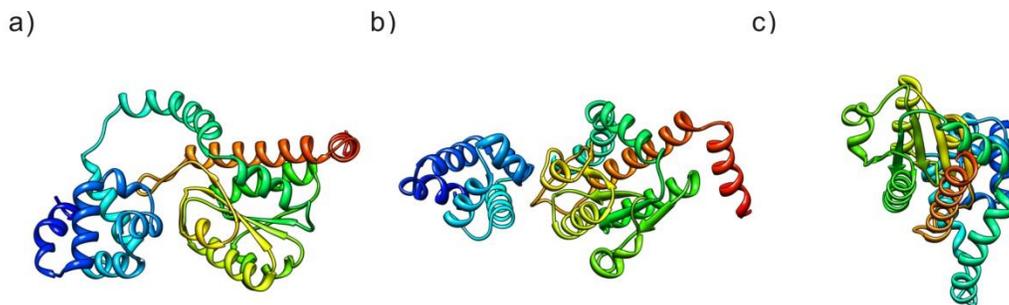


Figure 1.5 Predicted model of RpiRa structure. a) front, b) top, and c) side views of the predicted model of the entire RpiRa structure. The RpiRa homology model was determined using PSIPRED (<http://bioinf.cs.ucl.ac.uk/psipred/>) based on the *Vibrio vulnificus* NanR protein (PDB ID: 4IVN).⁸⁷

1.6 Overview of thesis

The study of metabolomics is making an increasingly beneficial impact on basic and applied biology. The challenge of metabolomics is to combine well-established instrumentation with novel biological experimental design. The desirable outcome of this effort is to yield a meaningful and reliable dataset to support a hypothesis or to inspire a new hypothesis.^{88,89} This dissertation describes my efforts to improve methodology critical to the successful application of NMR-based metabolomics and the application of these techniques to investigate the metabolome *S. epidermidis* and *S. aureus*.

In Chapter 2, a literature review summarizes the current progress in NMR-based metabolomics methodologies as applied to biofilm formation. This review demonstrates

the value and impact that NMR-metabolomics has had on the study of biofilms and exemplifies how metabolomics can assist in solving real-world problems. This review includes a detailed description of the history and current understanding of biofilm formation, the mechanism of molecular regulations under harsh environmental stressors, and many methods utilized to explore biofilms in a systems biology approach. In summary, it provides a practical and broad view of how to incorporate metabolomics into a biochemical research program.

The questions we asked ourselves everyday are how NMR spectroscopy is useful in metabolomics and how can we modify the current 1D and 2D NMR spectroscopic protocols so that they can provide more biologically relevant information. Thus, one important focus of my dissertation, which has had a significant impact on the field, is the establishment of protocols for NMR-based metabolomics. Dr. Steve Halouska made significant contributions to this project. A detailed description of our NMR-based metabolomics methodology is presented in Chapter 3. In Chapter 4, an in-house designed, high-throughput NMR tube washer is discussed. The NMR tube washer is used on routine and daily basis by a number of colleagues and undergraduate students.

Chapter 5-8 include two first- and two second-authored papers, showing the contribution of our NMR metabolomics method to investigate the metabolome from staphylococcal bacteria with our collaborator Dr. Somerville. The bacteria of *S. aureus* may cause a multitude of serious and acute infections, while infections with *S. epidermidis* are usually chronic and often involve the colonization and infection of indwelling medical devices.⁹⁰

In Chapter 5, a series of conditions that induce biofilm formation are examined and their effects on the TCA cycle are evaluated using NMR metabolomics. A relationship between biofilm formation and TCA cycle metabolism is established and well-characterized. Furthermore in Chapter 6, several environmental stressors are shown to depress the TCA cycle activity in *S. epidermidis* and cause the up-regulation of polysaccharide intracellular adhesion (PIA) biosynthesis. In Chapter 7, another metabolomics study is focused on the redox status of *S. aureus*. Iron limitation and aeration limitation are linked by oxidative stress and can lead to cell death. Fe or O₂ limitation can also trigger biofilm formation. Therefore, the impact of Fe and O₂ limitation on the *S. aureus* metabolome was investigated by NMR. A specific set of metabolites and metabolic pathways were identified that explain the physiological response of *S. aureus* to Fe and O₂ limitation. In Chapter 8, an *S. aureus* carbon catabolite repression (CcpE) deletion mutant, a postulated regulator of the TCA cycle, was constructed, and its metabolome was analyzed by NMR. Taken together, these studies strengthen our long-term hypothesis, originally postulated by Dr. Somerville, that the TCA cycle functions as a global regulator of many pathological processes in *S. epidermidis* and *S. aureus*. PIA, an important precursor of biofilm formation, is regulated by the TCA cycle.

In Chapter 9, one of the hypothetical regulatory proteins in the pentose phosphate pathway, the overexpression and multiple isotope labeling strategies for RpiRa is performed. The domain structure of this protein was predicted by bioinformatics, and the CTD is expected to be a ligand binding domain. The ligand-binding domain is predicted

to sense, an as yet, undefined metabolite to regulated PPP. RpiRa has been shown to be a PPP activator and can direct carbon flow away from the TCA cycle. Because of the size of the protein and NMR size-limitations, only a CTD plasmid was constructed and over-expressed in *E. coli*. The optimal condition and procedures for protein expression and purification is discussed. Preliminary studies in the structural determination are also addressed. Chapter 10 is a concluding chapter containing the summary of the work and future directions.

1.7 References

- (1) Baker, M. *Nat Meth* **2011**, 8, 117.
- (2) Suhre, K.; Shin, S.-Y.; Petersen, A.-K.; Mohny, R. P.; Meredith, D.; Wagele, B.; Altmaier, E.; CardioGram; Deloukas, P.; Erdmann, J.; Grundberg, E.; Hammond, C. J.; de Angelis, M. H.; Kastenmuller, G.; Kottgen, A.; Kronenberg, F.; Mangino, M.; Meisinger, C.; Meitinger, T.; Mewes, H.-W.; Milburn, M. V.; Prehn, C.; Raffler, J.; Ried, J. S.; Romisch-Margl, W.; Samani, N. J.; Small, K. S.; Erich Wichmann, H.; Zhai, G.; Illig, T.; Spector, T. D.; Adamski, J.; Soranzo, N.; Gieger, C. *Nature* **2011**, 477, 54.
- (3) Barupal, D. K.; Haldiya, P. K.; Wohlgemuth, G.; Kind, T.; Kothari, S. L.; Pinkerton, K. E.; Fiehn, O. *BMC Bioinformatics* **2012**, 13, 99.
- (4) Weljie, A. M.; Newton, J.; Mercier, P.; Carlson, E.; Slupsky, C. M. *Anal Chem* **2006**, 78, 4430.
- (5) Zhang, B.; Powers, R. *Future Med Chem* **2012**, 4, 1273.
- (6) Mailloux, R. J.; B ériault, R.; Lemire, J.; Singh, R.; Ch énier, D. R.; Hamel, R. D.; Appanna, V. D. *PLoS One* **2007**, 2, e690.

- (7) Somerville, G. A.; Proctor, R. A. *Microbiol Mol Biol Rev* **2009**, *73*, 233.
- (8) Roth, A.; Breaker, R. R. *Annu Rev Biochem* **2009**, *78*, 305.
- (9) Nicholson, J. K.; Lindon, J. C.; Holmes, E. *Xenobiotica* **1999**, *29*, 1181.
- (10) Bollard, M. E.; Contel, N. R.; Ebbels, T. M.; Smith, L.; Beckonert, O.; Cantor, G. H.; Lehman-McKeeman, L.; Holmes, E. C.; Lindon, J. C.; Nicholson, J. K.; Keun, H. C. *J Proteome Res* **2010**, *9*, 59.
- (11) Jones, O. A.; Spurgeon, D. J.; Svendsen, C.; Griffin, J. L. *Chemosphere* **2008**, *71*, 601.
- (12) Nicholson, J. K.; Lindon, J. C. *Nature* **2008**, *455*, 1054.
- (13) Garrod, S.; Humpfer, E.; Spraul, M.; Connor, S. C.; Polley, S.; Connelly, J.; Lindon, J. C.; Nicholson, J. K.; Holmes, E. *Magn Reson Med* **1999**, *41*, 1108.
- (14) Fiehn, O. *Plant Mol Biol* **2002**, *48*, 155.
- (15) Beloqui, A.; Guazzaroni, M. E.; Pazos, F.; Vieites, J. M.; Godoy, M.; Golyshina, O. V.; Chernikova, T. N.; Waliczek, A.; Silva-Rocha, R.; Al-Ramahi, Y.; La Cono, V.; Mendez, C.; Salas, J. A.; Solano, R.; Yakimov, M. M.; Timmis, K. N.; Golyshin, P. N.; Ferrer, M. *Science* **2009**, *326*, 252.
- (16) Oliver, S. G. *Nature* **1996**, *379*, 597.
- (17) Kind, T.; Fiehn, O. *Bioanal Rev* **2010**, *2*, 23.
- (18) Taylor, P. J. *Clinical Biochemistry* **2005**, *38*, 328.
- (19) Koal, T.; Deigner, H. P. *Curr Mol Med* **2010**, *10*, 216.

- (20) Aranibar, N.; Borys, M.; Mackin, N. A.; Ly, V.; Abu-Absi, N.; Abu-Absi, S.; Niemitz, M.; Schilling, B.; Li, Z. J.; Brock, B.; Russell, R. J., 2nd; Tymiak, A.; Reily, M. D. *J Biomol NMR* **2011**, *49*, 195.
- (21) Bollard, M. E.; Stanley, E. G.; Lindon, J. C.; Nicholson, J. K.; Holmes, E. *NMR Biomed* **2005**, *18*, 143.
- (22) León, Z.; García-Cañaveras, J. C.; Donato, M. T.; Lahoz, A. *ELECTROPHORESIS* **2013**, n/a.
- (23) Kim, H. K.; Verpoorte, R. *Phytochem Anal* **2010**, *21*, 4.
- (24) Bouatra, S.; Aziat, F.; Mandal, R.; Guo, A. C.; Wilson, M. R.; Knox, C.; Bjorn Dahl, T. C.; Krishnamurthy, R.; Saleem, F.; Liu, P.; Dame, Z. T.; Poelzer, J.; Huynh, J.; Yallou, F. S.; Psychogios, N.; Dong, E.; Bogumil, R.; Roehring, C.; Wishart, D. S. *PLoS One* **2013**, *8*, e73076.
- (25) Guo, A. C.; Jewison, T.; Wilson, M.; Liu, Y.; Knox, C.; Djoumbou, Y.; Lo, P.; Mandal, R.; Krishnamurthy, R.; Wishart, D. S. *Nucleic Acids Res* **2013**, *41*, D625.
- (26) Dunn, W. B.; Broadhurst, D.; Begley, P.; Zelena, E.; Francis-McIntyre, S.; Anderson, N.; Brown, M.; Knowles, J. D.; Halsall, A.; Haselden, J. N.; Nicholls, A. W.; Wilson, I. D.; Kell, D. B.; Goodacre, R. *Nat. Protocols* **2011**, *6*, 1060.
- (27) Hardy, N.; Taylor, C. *Metabolomics* **2007**, *3*, 243.
- (28) Castle, A. L.; Fiehn, O.; Kaddurah-Daouk, R.; Lindon, J. C. *Brief Bioinform* **2006**, *7*, 159.
- (29) Powers, R. *Journal of Medicinal Chemistry* **2014**.
- (30) Zhang, B.; Halouska, S.; Schiaffo, C. E.; Sadykov, M. R.; Somerville, G. A.; Powers, R. *J Proteome Res* **2011**, *10*, 3743.
- (31) Halouska, S.; Fenton, R. J.; Barletta, R. G.; Powers, R. *ACS Chem Biol* **2012**, *7*, 166.

- (32) Chaika, N. V.; Gebregiworgis, T.; Lewallen, M. E.; Purohit, V.; Radhakrishnan, P.; Liu, X.; Zhang, B.; Mehla, K.; Brown, R. B.; Caffrey, T.; Yu, F.; Johnson, K. R.; Powers, R.; Hollingsworth, M. A.; Singh, P. K. *Proc Natl Acad Sci U S A* **2012**, *109*, 13787.
- (33) Riley, M. *Microbiol Rev* **1993**, *57*, 862.
- (34) Duarte, N. C.; Becker, S. A.; Jamshidi, N.; Thiele, I.; Mo, M. L.; Vo, T. D.; Srivas, R.; Palsson, B. O. *Proc Natl Acad Sci U S A* **2007**, *104*, 1777.
- (35) Heinemann, M.; Kummel, A.; Ruinatscha, R.; Panke, S. *Biotechnol Bioeng* **2005**, *92*, 850.
- (36) Richardson, D. J. *Microbiology* **2000**, *146* (Pt 3), 551.
- (37) Plugge, C. M.; Zhang, W.; Scholten, J. C.; Stams, A. J. *Front Microbiol* **2011**, *2*, 81.
- (38) Lu, Y.-K.; Marden, J.; Han, M.; Swingley, W.; Mastrian, S.; Chowdhury, S.; Hao, J.; Helmy, T.; Kim, S.; Kurdoglu, A.; Matthies, H.; Rollo, D.; Stothard, P.; Blankenship, R.; Bauer, C.; Touchman, J. *BMC Genomics* **2010**, *11*, 325.
- (39) Freilich, S.; Kreimer, A.; Borenstein, E.; Yosef, N.; Sharan, R.; Gophna, U.; Ruppin, E. *Genome Biology* **2009**, *10*, R61.
- (40) Raamsdonk, L. M.; Teusink, B.; Broadhurst, D.; Zhang, N.; Hayes, A.; Walsh, M. C.; Berden, J. A.; Brindle, K. M.; Kell, D. B.; Rowland, J. J.; Westerhoff, H. V.; van Dam, K.; Oliver, S. G. *Nat Biotechnol* **2001**, *19*, 45.
- (41) Wishart, D. S. *Brief Bioinform* **2007**, *8*, 279.
- (42) Kauffman, K. J.; Prakash, P.; Edwards, J. S. *Curr Opin Biotechnol* **2003**, *14*, 491.
- (43) Fell, D. *Understanding the Control of Metabolism*; Portland Press, London: Portland, 1997.

- (44) Varma, A.; Palsson, B. O. *Appl Environ Microbiol* **1994**, *60*, 3724.
- (45) Grafahrend-Belau, E.; Junker, A.; Eschenröder, A.; Müller, J.; Schreiber, F.; Junker, B. H. *Plant Physiology* **2013**.
- (46) Gu, H.; Gowda, G. A.; Raftery, D. *Future Oncol* **2012**, *8*, 1207.
- (47) Worley, B.; Halouska, S.; Powers, R. *Anal Biochem* **2013**, *433*, 102.
- (48) Werth, M. T.; Halouska, S.; Shortridge, M. D.; Zhang, B.; Powers, R. *Anal Biochem* **2010**, *399*, 58.
- (49) Wei, S.; Zhang, J.; Liu, L.; Ye, T.; Gowda, G. A.; Tayyari, F.; Raftery, D. *Anal Chem* **2011**, *83*, 7616.
- (50) Gu, H.; Gowda, G. A. N.; Neto, F. C.; Opp, M. R.; Raftery, D. *Analytical Chemistry* **2013**, *85*, 10771.
- (51) Cloarec, O.; Dumas, M. E.; Craig, A.; Barton, R. H.; Trygg, J.; Hudson, J.; Blancher, C.; Gauguier, D.; Lindon, J. C.; Holmes, E.; Nicholson, J. *Anal Chem* **2005**, *77*, 1282.
- (52) Halouska, S.; Powers, R. *J Magn Reson* **2006**, *178*, 88.
- (53) Hu, K.; Westler, W. M.; Markley, J. L. *J Am Chem Soc* **2011**, *133*, 1662.
- (54) Bingol, K.; Zhang, F.; Bruschiweiler-Li, L.; Br üschweiler, R. *Analytical Chemistry* **2013**, *85*, 6414.
- (55) Tayyari, F.; Gowda, G. A.; Gu, H.; Raftery, D. *Anal Chem* **2013**, *85*, 8715.
- (56) Kikuchi, J.; Shinozaki, K.; Hirayama, T. *Plant Cell Physiol* **2004**, *45*, 1099.
- (57) Kikuchi, J.; Hirayama, T. *Methods Mol Biol* **2007**, *358*, 273.

- (58) O'Toole, G.; Kaplan, H. B.; Kolter, R. *Annu. Rev. Microbiol.* **2000**, *54*, 49.
- (59) Costerton, J. W.; Lewandowski, Z.; Caldwell, D. E.; Korber, D. R.; Lappin-Scott, H. M. *Annu Rev Microbiol* **1995**, *49*, 711.
- (60) Costerton, J. W. *J Ind Microbiol* **1995**, *15*, 137.
- (61) Angst, E., C.; Navy, U., Ed. Washington, DC, 1923.
- (62) Henrici, A., T., *Journal of Bacteriology* **1933**, *1933*, 10.
- (63) Heukelekian, H., Heller, A., *J Bacteriol.* **1940**, *40*, 12.
- (64) Hiroyuki, K. *The American Journal of Medicine* **1995**, *99*, 26s.
- (65) Klaus-Joerger, T.; Joerger, R.; Olsson, E.; Granqvist, C.-G. *Trends in Biotechnology* **2001**, *19*, 15.
- (66) van Loosdrecht, M. C. M.; Eikelboom, D.; Gjaltema, A.; Mulder, A.; Tjihuis, L.; Heijnen, J. J. *Water Science and Technology* **1995**, *32*, 35.
- (67) *Biofilms in Medicine, Industry and Environmental Biotechnology- Characteristics, Analysis and Control*; IWA Publishing London, 2003.
- (68) Gannon, J., C., Washington, DC, 2000.
- (69) Lewis, K. *Antimicrob Agents Chemother* **2001**, *45*, 999.
- (70) Davies, D. *Nat Rev Drug Discov* **2003**, *2*, 114.
- (71) Kennedy, A. D.; Otto, M.; Braughton, K. R.; Whitney, A. R.; Chen, L.; Mathema, B.; Mediavilla, J. R.; Byrne, K. A.; Parkins, L. D.; Tenover, F. C.; Kreiswirth, B. N.; Musser, J. M.; DeLeo, F. R. *Proc Natl Acad Sci U S A* **2008**, *105*, 1327.

- (72) Nguyen, D.; Joshi-Datar, A.; Lepine, F.; Bauerle, E.; Olakanmi, O.; Beer, K.; McKay, G.; Siehnel, R.; Schafhauser, J.; Wang, Y.; Britigan, B. E.; Singh, P. K. *Science* **2011**, *334*, 982.
- (73) O'Toole, G.; Kaplan, H. B.; Kolter, R. *Annu Rev Microbiol* **2000**, *54*, 49.
- (74) Fung, D. K.; Chan, E. W.; Chin, M. L.; Chan, R. C. *Antimicrob Agents Chemother* **2010**, *54*, 1082.
- (75) Antunes, L. C.; Ferreira, R. B. *Crit Rev Microbiol* **2009**, *35*, 69.
- (76) Yu, D.; Zhao, L.; Xue, T.; Sun, B. *BMC Microbiol* **2012**, *12*, 288.
- (77) Sadykov, M. R.; Zhang, B.; Halouska, S.; Nelson, J. L.; Kreimer, L. W.; Zhu, Y.; Powers, R.; Somerville, G. A. *J Biol Chem* **2010**, *285*, 36616.
- (78) Wen, H.; Yang, H. J.; An, Y. J.; Kim, J. M.; Lee, D. H.; Jin, X.; Park, S. W.; Min, K. J.; Park, S. *Mol Cell Proteomics* **2013**, *12*, 575.
- (79) Frezza, C.; Zheng, L.; Tennant, D. A.; Papkovsky, D. B.; Hedley, B. A.; Kalna, G.; Watson, D. G.; Gottlieb, E. *PLoS One* **2011**, *6*, e24411.
- (80) Zhu, Y.; Nandakumar, R.; Sadykov, M. R.; Madayiputhiya, N.; Luong, T. T.; Gaupp, R.; Lee, C. Y.; Somerville, G. A. *J Bacteriol* **2011**, *193*, 6187.
- (81) Sorensen, K. I.; Hove-Jensen, B. *J Bacteriol* **1996**, *178*, 1003.
- (82) Pearson, W. R. *Methods Mol Biol* **2000**, *132*, 185.
- (83) Bryson, K.; Cozzetto, D.; Jones, D. T. *Curr Protein Pept Sci* **2007**, *8*, 181.
- (84) Quevillon, E.; Silventoinen, V.; Pillai, S.; Harte, N.; Mulder, N.; Apweiler, R.; Lopez, R. *Nucleic Acids Research* **2005**, *33*, W116.
- (85) Rost, B.; Yachdav, G.; Liu, J. *Nucleic Acids Res* **2004**, *32*, W321.

- (86) McGuffin, L. J.; Bryson, K.; Jones, D. T. *Bioinformatics* **2000**, *16*, 404.
- (87) Bryson, K.; McGuffin, L. J.; Marsden, R. L.; Ward, J. J.; Sodhi, J. S.; Jones, D. T. *Nucleic Acids Research* **2005**, *33*, W36.
- (88) Parab, G. S.; Rao, R.; Lakshminarayanan, S.; Bing, Y. V.; Mochhala, S. M.; Swarup, S. *Anal Chem* **2009**, *81*, 1315.
- (89) Reinke, S. N.; Broadhurst, D. I. *Genome Med* **2012**, *4*, 85.
- (90) Otto, M. *Nat Rev Micro* **2009**, *7*, 555.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Biofilms are a natural part of the ecology of the earth, and correspond to a “social structure” of microorganisms compared to a planktonic state.¹⁻³ Biofilms are a ubiquitous microbial phenomenon that has been observed for bacteria, fungi, algae, yeasts, protozoa, and other organisms. In a biofilm composed of multiple diverse organisms, the interspecies interaction can range from neutral, to cooperative, to competitive, and finally to antagonistic.⁴ The diversity of organisms able to self-organize and form biofilms is quite astounding and may provide clues to the evolution of multicellular organisms.⁵ Are biofilms a transitional state of evolution and the basis for multicellular organisms? Or are biofilms simply a highly organized state of single cell organisms? The fact that biofilms provide a significant survival advantage for adapting to the harsh and distinct environmental conditions probably explains its broad adaptation.

Biofilms have been studied for almost a century. The fouling of ships’ hulls by marine microorganisms was documented in 1923.⁶ In 1933, Henrici observed “it is quite evident that for the most part water bacteria are not free floating organisms, but grow upon submerged surfaces”.⁷ In 1940, Heukelekian and Heller added "development takes place either as bacterial slime or colonial growth attached to surfaces".⁸ However, it was not until the last decade of the last century that significant progress was made to effectively study microbial communities. It was then that we began to appreciate the significant implications of biofilms. Nevertheless, there are still many gaps in our

understanding of the biological processes related to this phenomenon.^{1,9-11}

The interest in biofilm is not merely a result of scientific curiosity, but is also derived from practical concerns related to medical science,¹² material engineering,¹³ civil engineering,¹⁴ and others.¹⁵ In the area of medicine, research on biofilms has focused on its relationship to bacterial infections and drug resistance. Bacterial infections are a serious disease and major source of world-wide deaths. Especially concerning is the growing resistance to antibiotics that has become a major medical issue in developing countries. Between 1980 and 1992, infectious disease deaths increased by 58%; the major contributors were HIV infections and AIDS, respiratory disease, and bloodstream infection.¹⁶ A 2000 US government report identified infectious diseases as a leading cause of death worldwide and the third leading cause of death in the United States.¹⁷ It has been estimated that 60%-80% of human microbial infections are caused by bacteria growing as a biofilm.¹⁸ Certain pathogenic biofilms are particularly concerning because of the added issue of drug-resistance.¹⁹ Multidrug resistant pathogens, such as *Enterococcus faecium*, *Klebsiella pneumonia*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter* species are infecting the majority of US hospitals.²⁰ As methicillin-resistant *Staphylococcus aureus* (MRSA) and other resistant pathogens capable of biofilm formation continue to emerge and propagate, understanding and circumventing biofilm resistance to antibiotics is a paramount necessity.²¹

Pathogens can be introduced into the human body through trauma, medical operations, dental procedures, or by other means.²²⁻²⁴ Many surfaces of organs are heavily colonized by microbes that have the potential to cause an infection, especially

during any invasive medical procedure. In fact, the ratio of bacteria to mammalian cells living within the human body is ten to one, providing ample opportunity for inducing a bacterial infection from medical procedures or trauma.²⁵ For example, there are over 500 species of microorganisms identified in typical dental plaque.²⁶ Correspondingly, dental cavities (caries) are commonly a result of bacterial biofilm infections.²⁷ Biofilms are also formed on our tongues, cheeks, in our intestines, nasal passages, sinuses, and on our skin.²⁶ These human microbial communities are largely unstudied and their role in infections is largely unknown. But biofilms protect the organisms from both antimicrobials and the host immune response, making infectious biofilms extremely difficult to treat.²⁸ For instance, staphylococcal biofilm infections have a 10 to 1000 fold increase in antibiotic resistance.²⁹⁻³¹

A serious source of biofilm infections is heart disease, which is also a major cause of mortality in the US.³² Invasive surgical techniques are inevitably required to treat the resulting symptoms of heart disease, which may lead to fatal staphylococcal infective endocarditis (IE). *S. aureus* and *S. epidermidis*³³ infections stemming from implantable medical devices (*e.g.*, pace makers,^{34,35} indwelling vascular catheters,³⁶⁻³⁸ grafts,³⁹ and left ventricular assist devices⁴⁰⁻⁴⁴) are common causes of IE. Biofilms have also been identified on various other medical devices^{18,45} such as contact lenses, endotracheal tubes, central venous catheters, pacemakers, and voice prostheses that account for over 80% of microbial infections in the body. Catheter-associated urinary tract infection is also a common source of biofilm infections. It has also been suggested that autoimmune disorders like arthritis, chronic fatigue syndrome, fibromyalgia, Crohn's disease, and

ulcerative colitis are caused by biofilm infections.⁴⁶ In summary, bacterial biofilms pose a serious threat to human health because of the added protection biofilms provide from an immune response and antibiotic treatments, because of the ease of acquiring an infection from trauma and medical procedures, and because of the rapid emergence of drug resistance among bacteria that form biofilms.

2.2 Biofilm Structure

Initial scientific studies on bacterial biofilms have focused on understanding the structure and composition of biofilms (Figure 2.1). Confocal scanning laser microscopy (CSLM) combined with fluorescent molecular probes and antibody labeling has been used to provide invaluable information on biofilm cell morphology, cellular metabolism, and cell phylogeny.^{47,48} These studies revealed heterogeneous features for both pure-cultured and mix-cultured biofilms, and the adaptability of unicellular organisms to a diverse range of physical, chemical and communal circumstances.^{49,50} Basically, a biofilm is comprised of three parts: (1) a living or non-living substance that provides a moist surface for attachment of the highly-organized microbial structure,⁵¹⁻⁵³ (2) a slimy-like matrix made of extracellular DNA, proteins, and polysaccharides (β (1-6)-linked N-acetylglucosamine polymer)^{54,55} that embeds the microorganism,⁵⁶ and (3) an aggregate of microorganisms in a community that exchange fluids, nutrients and chemical signals.⁵⁷ Biofilms are viewed as layers of bacteria encapsulated within different microenvironments due to variations in nutrient availability⁵⁸ and differing cell densities.^{50,59} Bacteria within the biofilm core exist in a stationary or dormant growth

phase⁵⁸ and are physiologically distinct from planktonic bacteria.⁶⁰⁻⁶³ Antibiotics that target cellular mechanisms associated with growing bacteria have diminished activity against biofilms.^{64,65} Presumably, the antibiotic is only affective against the fast growing bacteria on the biofilm's surface. The dormant bacteria in the interior can replenish a biofilm's outer layers after the antibiotic has been withdrawn. In order for these dormant bacteria to enter a growth phase, it is necessary that they respond to a change in nutrient and cell density resulting from the death of the fast growing bacteria on the surface.

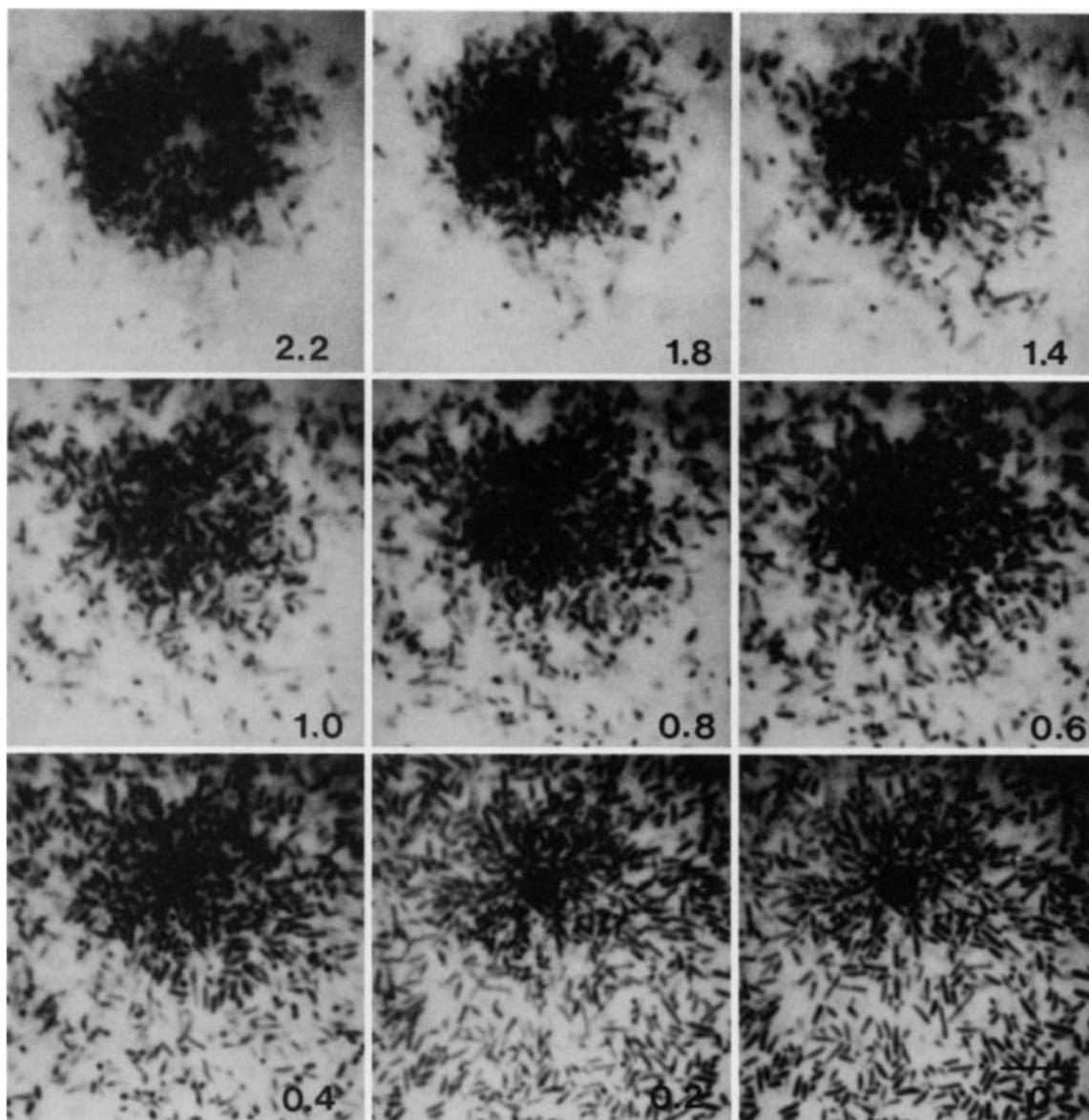


Figure 2.1 Horizontal optical thin sections (0 to 2.6 μm) of the *P. aeruginosa* biofilm obtained by SCLM. The biofilm was negatively stained with 0.1% fluorescein. The horizontal sections show the removal of out-of-focus information and reveal aspects of the internal structure of the biofilm. Bar=5 μm . Reprinted with permission from reference.⁵⁰

Bacterial cells have also been shown to adopt various functions critical to the

biofilm structure, such as forming the initial attachment point with the surface, producing the slime matrix, forming water channels to deliver oxygen and other nutrients, or comprising the simple building blocks of the biofilm.² Bacterial cells within a micro-colony can maintain homeostasis, achieve flexible spatial relationships with cooperative organisms, and create an effective way of exchanging nutrients and metabolites with the environment. The realization that biofilms consist of such a complex structure raised intriguing scientific questions: how do bacteria create such an elaborate architecture and what stimulates biofilm formation?

The life cycle of the biofilm turns out to be even more complicated and can be divided into roughly three steps: attachment, growth and propagation (Figure 2.2). First, a few colonies reversibly adhere to the surface via van der Waals forces to create an initiation site. Attachment involves lipoteichoic acid anchored to the cell membrane.⁵⁷ This is followed by an irreversible attachment of the cells through the production of the exopolysaccharide (EPS) matrix and cell growth.⁵² The cell growth is not uniform and results in the formation of channels.⁹ A combination of cell division and recruitment occurs during the maturation stage and only biofilm shape and size are changed.^{9,53,66} Other organisms may be recruited to the biofilm depending on the microenvironments created by on-going metabolic activity. Finally, detachment of individual cells and dispersion enables the biofilm to spread and colonize new surfaces or to join another biofilm.^{67,68} Biofilms form irregular spatial structures, which are affected by many different ecological, biological, chemical, and physical factors. The effect of these factors on biofilm formations have been investigated using a variety of computer simulations.⁶⁹⁻

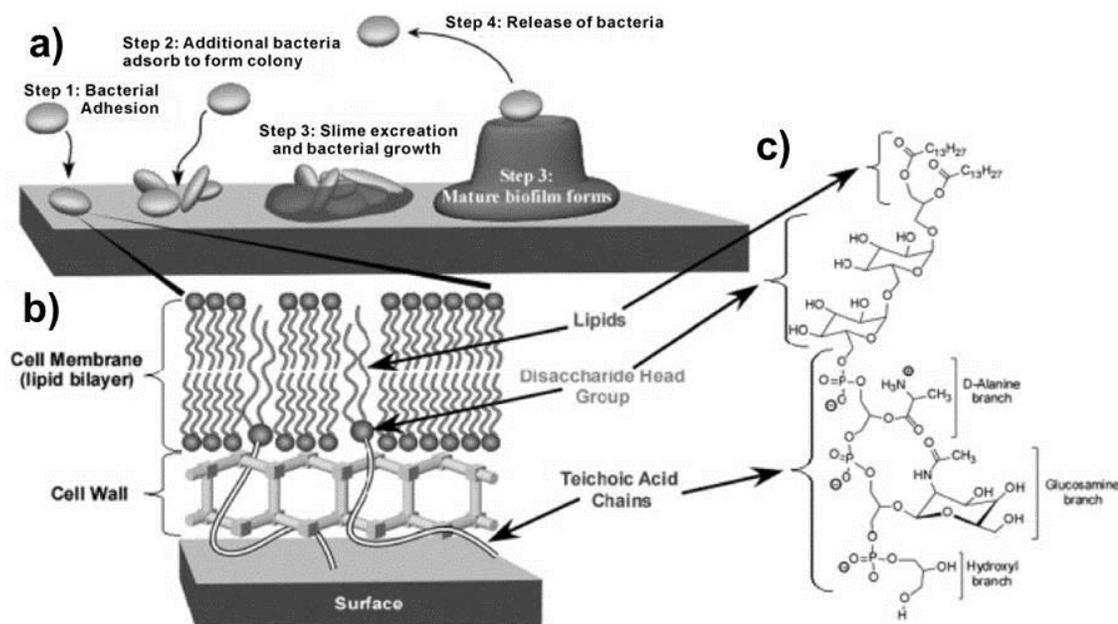


Figure 2.2 a) The four-step process involved in the formation of biofilms. b) The lipoteichoic acid is anchored to the cell membrane and extends out past the edge of the cell wall being the initial molecular contact between the bacterium and other materials. c) The chemical structure of lipoteichoic acid. Reprinted with permission from reference.⁵⁷

Biofilm models have been reported in the literature for different organisms and biofilms composed of multispecies, for different nutritional conditions, and for different environments. These mathematical models have been used to investigate the interrelationship of various biofilm factors, especially for parameters that can be difficult to vary experimentally.⁷⁴ There are three primary biofilm mechanisms: (i) substrate transport, (ii) growth, and (iii) necrosis and loss, which are applied to any computational model.^{75,76} Biofilm models can be grouped into four different categories:⁷⁷ (1) one-

dimensional continuum models, (2) diffusion limited aggregation models, (3) cellular automaton models, and (4) fully coupled biofilm-fluid models. One-dimensional continuum models assume gradients are orders of magnitude higher in the direction perpendicular to the attachment surface. Diffusion limited aggregation models have an immobile seed located on a square lattice. One particle after another is released from a random location from the seed and is allowed to diffuse through a random walk to form an aggregate when it contacts the seed particle.⁷⁸ Cellular automaton models consist of a regular grid of cells where each cell is in one of a finite number of states. Rules determine cell movement, replication, death, *etc.* as a function of micro-environmental changes. As an example, different morphologies from dense layers to open, mushroom-like forms can be simulated. Effectively, small changes in nutrient concentration can have a dramatic effect on biofilm development.⁷⁹ Fully coupled biofilm-fluid models are more complex simulations and also take into account the hydrodynamics of the bulk fluid and the non-uniform distribution of cells and polymers.⁷⁷

2.3 Biofilm formation at the molecular level

There are some general principals regarding biofilm formation that include the need for metabolically active bacteria for surface adhesion, the need for an adequate nutrient supply for cell replication and exopolysaccharide production, and the fact that surfaces coated with organic nutrients stimulate biofilm formation.^{2,3} Correspondingly, bacteria biofilms readily form on the surfaces of plastic or metal medical devices in body fluids. Some common bacterial biofilm infections include *Pseudomonas aeruginosa* in

the lung, *Escherichia coli* in the urinary tract, *Vibrio cholera* in the gastrointestinal tract, *S. epidermidis* in the heart, *S. aureus* in arteries, Enterococcus species in the urinary tract, and fungi such as *Candida* species in gastrointestinal tract.⁸⁰ The identification of biochemical pathways critical to biofilm formation is an important first-step to being able to prevent these bacterial biofilm infections. Even with our general understanding of the basic structure and development of bacterial biofilms, comprehending the underlying processes responsible for inducing the transition from planktonic cells to a biofilm is still unclear. Correspondingly, the planktonic to biofilm transition is a complex and highly regulated process that results in a phenotypic change. Thus, the differential expression and regulation of specific genes are associated with biofilm formation.

Genomics analysis of biofilm formation started in the 1990s by first screening for biofilm-defective mutants.^{81,82} Such efforts identified a diverse number of genes required for biofilm formation.⁸³⁻⁸⁹ More recently, DNA microarray technology has been used to identify genes up- or down-regulated in bacterial biofilms.⁹⁰ Unfortunately, there does not appear to be a clear trend in biofilm related genes. Instead, multiple pathways to biofilm formation that depend on media, growth conditions, and specific organism is likely.⁹¹ Nevertheless, some broad, common features have been observed, such as the up-regulation of genes for polysaccharide production, for various stress-induced pathways, for stationary-phase induced genes, for a prevalence of genes of unknown function, and new regulatory pathways.⁹⁰ For example, Quoc *et al.* (2007) identified nineteen genes in *S. aureus* associated with biofilm formation that were not previously observed.⁸⁷ Again, this highlights the difficulty encountered with identifying a uniform set of biofilm-related

genes. Besides genes involved in polysaccharide intercellular adhesion (PIA) or unknown function, the authors observe mutations in guanosine-dependent regulation and formation of wall teichoic acid. The complexity of biofilm formation may be attributed to the fact that the transition from planktonic cells to a biofilm is influenced by various and diverse environmental factors such as ethanol,⁹² oleic acid,⁹³ glucose,⁹⁴ UDP-N-acetylglucosamine,⁹⁵ sub-inhibitory concentrations of some antibiotics,⁹⁶ anaerobic conditions,⁹⁷ Fe limitation,⁹⁸⁻¹⁰⁰ high osmolarity,¹⁰¹ and high temperature.¹⁰¹ The diversity of these external stimuli suggests a versatile regulation system.

After two decades of research, multiple mechanisms of biofilm regulation have been proposed: sigma factors, two-component systems, or quorum sensing. Sigma factors control the expression of various genes including virulence factors and global regulators that are related to biofilm formation.¹⁰²⁻¹⁰⁴ Sigma factors are activated when bacteria sense environmental conditions that induce stress (heat shock, nitrogen-limitation, starvation, high osmolarity, *etc.*).¹⁰⁵ Two-component systems are an alternative stimulus-response coupling mechanism that have been shown to regulate diverse metabolic processes such as the bacterial cell cycle, cell-cell communication, and virulence factors in biofilm formation.¹⁰⁶ In a two-component system, a ligand or a signal molecule can stimulate a histidine kinase sensor protein, which undergoes autophosphorylation at a conserved histidine residue. The phosphoryl group is then transferred to the cognate response regulator, which can activate or repress transcription of the target genes.^{106,107} Conversely, quorum sensing uses signal molecules for bacterial intercellular communication. Quorum sensing enable bacteria to “sense” cell density and coordinate

behavior in response to nutrient availability, toxic compounds, host-immune response, and defense.^{108,109} In gram-negative bacteria, N-acyl homoserine lactones (autoinducer-1, AI-1) have been identified as the signal molecules.¹¹⁰ AI-1 is synthesized and sensed by analogous LuxI and LuxR regulatory proteins. The specific AI-1 molecule varies between gram-negative organisms. For gram-positive bacteria, autoinducer peptides (with no conserved sequence) have been identified as a signal molecule that involves a two-component signal transduction system.¹¹¹ Furanosyl borate diester (AI-2) has been identified as a universal interspecies signal molecule that regulates biofilm formation in over 55 gram-positive and gram-negative species.¹¹² Figure 2.3 illustrates some common regulatory mechanisms of the planktonic to biofilm transition.

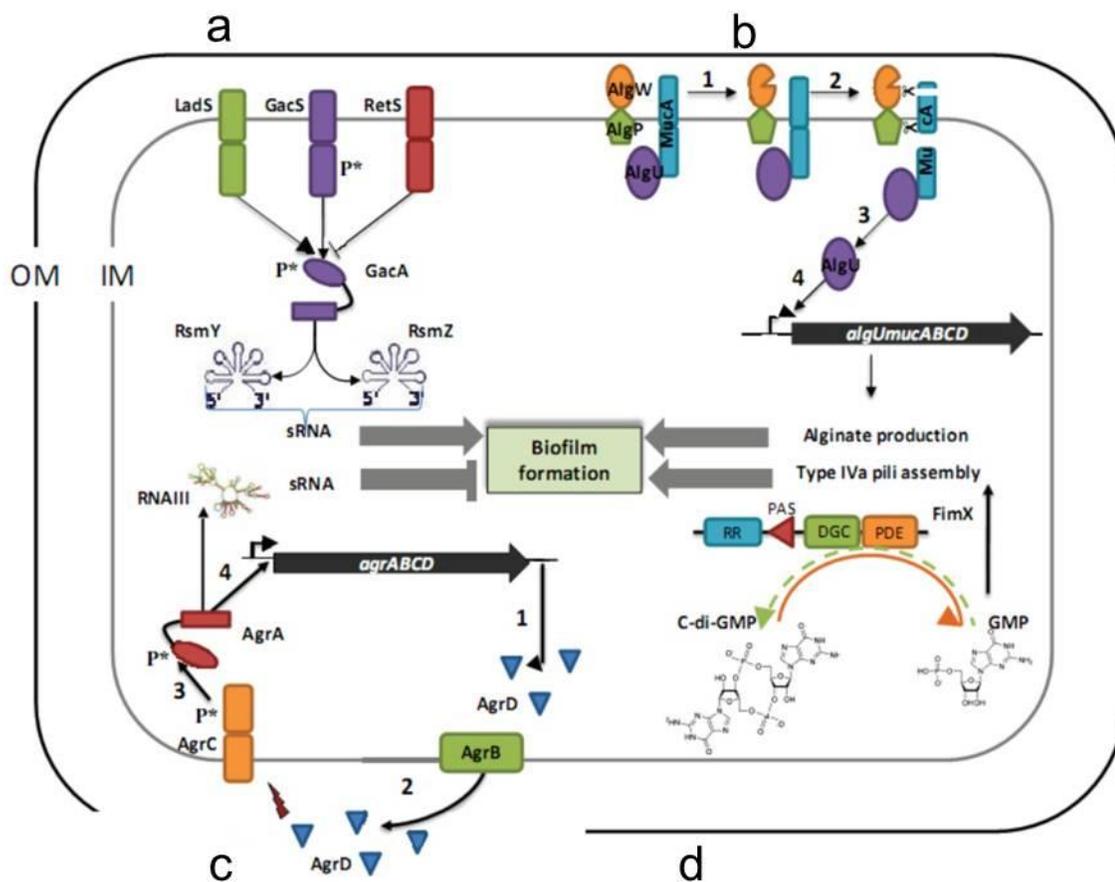


Figure 2.3 Regulatory networks controlling transition between planktonic and biofilm lifestyle. The external frames illustrate the bacterial envelope with one or two membranes (OM: outer membrane, IM: inner membrane) according to Gram-positive (c) and Gram-negative bacteria (a, b, and d), respectively. a) Control of biofilm formation in *P. aeruginosa* through the TCS GacS (HK)/GacA (RR) mediated by sRNA *rsmY* and *rsmZ* gene transcription and modulated by RetS and LadS, two additional HK in *P. aeruginosa*. b) Control of EPS alginate in *P. aeruginosa*, which further impacts biofilm architecture by the system ECF sigma factor AlgU - anti-sigma MucA - AlgP (IM)-AlgW (periplasmic) complex: 1) activation of AlgW/AlgP, 2) cleavage of MucA, 3) release of AlgU, 4) activation of the *alg UmucABCD* operon. c) Control of *S. aureus* biofilm formation through the Agr QS system: 1) AgrD autoinducer production, 2) AgrD autoinducer accumulation in the extracellular medium where it reaches a threshold, 3)

activation of the TCS AgrCA by AgrD at the threshold concentration, 4) AgrA-dependent activation of the sRNA RNA III expression repressing expression of genes involved in biofilm formation together with amplification loop of *agrABCD*. d) Control of *P. aeruginosa* biofilm formation through the intracellular second messenger c-di-GMP level controlled by the FimX protein having DGC and PDE domains, a RR domain, and a PAS domain. Note that in FimX protein only PDE activity is detectable (continuous arrow), whereas DGC activity is undetectable (dotted arrow). Reprinted with permission from reference.¹¹³

Two-component systems (TCS) and autoinducers are promising drug targets for biofilms because of the essential role in cell growth regulation and the unique mechanisms of action compared to conventional antibiotics.¹¹⁴⁻¹¹⁷ Importantly, proteins from TCS and quorum sensing are absent in humans, minimizing toxicity concerns. Also, both biological processes are based on ligand-receptor interactions, which are typical targets for drug discovery and have a reasonable likelihood of success. Histidine kinases and response regulators, such as WalK/WalR, YhcS/YhcR81, HP165/HP166, and MtrB/MtrA are potential drug targets for bacteria pathogens.¹¹⁷ Targeting the kinase domain appears to suffer from poor selectivity, but targeting the sensor domain may prove more successful. Also, targeting non-essential TCS proteins that regulate virulence such as GacS/GacA, PhoQ/PhoP, and CorS/CorR, has demonstrated some initial positive results.¹¹⁷

Alternatively, quorum sensing may be interrupted by targeting the LuxI, LuxR, or LuxS transcriptional regulators,^{118,119} AIP receptors,¹¹⁶ or Lsr transporters.¹¹² A common approach is to use the three classes of autoinducers as chemical templates to design

agonists or antagonists as a starting-point for drug design (Figure 2.4).^{120,121} For instance, TCS proteins QseC/QseB from *Escherichiacoli* (EHEC) O157:H7 that responds to AI-3, epinephrine, and norepinephrine, are inhibited by LED209 (*N*-phenyl-4-[[[(phenylamino)thioxomethyl]amino]-benzenesulfonamide]).¹²² Additionally, the RNAIII-inhibiting peptide (RIP; YSPWTNF-NH₂) is an inhibitor of the staphylococcal TRAP/*agr* system, which is regulated by autoinducer RNAIII-activating protein.¹²³ RIP was shown to be active against methicillin-resistant *Staphylococcus aureus* graft infections. There has also been significant effort in the design of AI-2 analogs as novel antibiotics because of AI-2's broad activity against multiple species.^{124,125} For instance Roy *et al.* (2010) explored the activity C-1 alkyl analogs of AI-2 against multiple bacterial species. Ethyl-DPD (DPD; 4,5-dihydroxy-2,3-pentanedione) was found to inhibit quorum-sensing in both *Escherichia coli* and *Salmonella typhimurium*. In addition, Rui *et al.* (2012) explored DPD analogs with a new stereocenter at C-5 (4*S*, 5*R*)-DHD.¹²⁶ The compound was also shown to be active against both *E. coli* and *Vibrio harveyi*. Conversely, Tsuchikama *et al.* (2011) synthesized carbocyclic analogues of DPD that were inactive against *S. typhimurium* and *V. harveyi*, suggesting the importance of the linear form of DPD and the heterocyclic oxygen atom.¹²⁷ Similarly, Lowery *et al.* (2005) explored a variety of DPD analogs that resulted in a uniform lower activity that highlights the general challenge in evolving a small chemical template into a drug.¹²⁸ Nevertheless, the unique mechanism of action for TCS and autoinducer targets holds the promise of a valuable new class of drugs that may circumvent biofilm resistance to antibiotics (Figure 2.4).

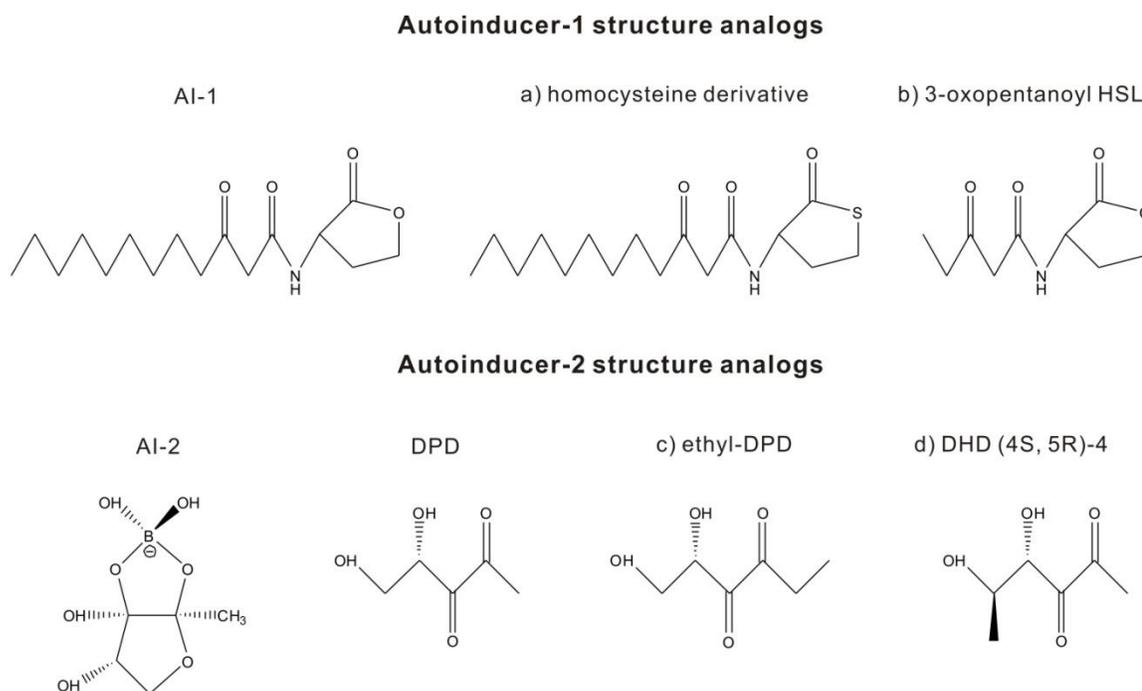


Figure 2.4 Structures of autoinducers and analogs. (top) AI-1 chemical structure and its analogs a) homocysteine derivative¹²⁰; b) 3-oxopentanoyl HSL¹²¹. (bottom) AI-2 chemical structure, its precursor DPD and its analogs c) ethyl-DPD¹²⁹; d) DHD (4S, 5R)-4.¹³⁰

2.4 An Overview of Metabolomics

Metabolomics is the study of metabolites, such as amino acids, carbohydrates and lipids that are the end products of cellular regulatory processes, as well as intermediates and other signaling molecules.^{131,132} The metabolome is the complete collection of all metabolites within a biological cell compartment, cell, tissue, organ or organism examined in the form of a cellular extract or biofluids.^{133,134} In general, a molecular weight of 1 kDa is the typical limit that separates metabolites from macromolecules.¹³⁵

There are many differences between conventional metabolite measurements and metabolomics. First, metabolomics focuses on a global or broad-based analysis of

metabolites through a high-throughput detection methodology compared to a limited and directed analysis of a specific number of individual metabolites.¹³² In general, metabolomics does not require the complete separation of individual metabolites. Instead, it captures a “snap-shot” or “fingerprint” of the state of the metabolome. Thus, metabolomics simplifies metabolite detection by using a single analytical technique to characterize the state of the metabolome. In this manner, metabolomics also provides an unbiased view of changes in metabolism by covering all major pathways. Thus, the systematic analysis of the ultimate response of a biological system has a better chance of describing pleiotropic effects.¹³² Second, metabolomics uses a combination of multiple methodologies, such as cellular biology, instrumental analysis, chemometrics, and bioinformatics to analyze the biological system. This combination of techniques provides a better view of the global role that metabolism plays in cellular functions. Again the analysis of a select set of metabolites does not provide this sort of global picture of cellular activity. Yet in theory, it should be possible to correlate metabolic changes in a biochemical pathway with the enzymes involved, and then to the underlying genetic alterations or changes in gene expression or regulation.¹³² A computational simulation could also integrate the experimental data to create a systematic view of the effected biochemical pathways, and potentially, the relevant proteins. The identification of specific proteins that are disease-related, or in this case related to biofilm formation, is a fundamental and critical step of the drug discovery process.

Autoinducers are an excellent example of the significant roles that small molecules can play in bacterial biofilms, where mimicking or inhibiting the mode of

action of autoinducers is a potential drug discovery target. Other inter- or intracellular small molecular-weight molecules may have similar roles in the initiation, progression and survivability of bacterial biofilms. Specifically, biofilms are a spatial distribution of heterogeneous cells, where cells exist in different metabolic states to maximize survival. Thus, understanding biofilms requires a comprehensive characterization of the various metabolic states within a complex cellular community. Correspondingly, metabolomics provides a systematic approach to explain this complex system.^{136,137}

Compared to genes and proteins, primary metabolites are highly conserved between various cells and organisms.¹³⁸ In a similar manner to gene expression, some of the key metabolites, nutrients, and signal molecules (autoinducers) have been shown to significantly influence biofilm formation.¹³⁹ Thus, metabolite quantification and the pathway modeling of complex biological systems is useful for exploring cell behavior in establishing a biofilm community. Furthermore, metabolomics is an invaluable approach for investigating antibiotic resistance in biofilms. By generating a network of metabolites affected by the drug treatment, it is possible to predict the antibiotic's mechanism of action.¹⁴⁰ Additionally, the phenotype of antibiotic resistance strains and biofilm strains can be characterized through their relative metabolome differences. Similarly, monitoring metabolic changes can be used to investigate the effects of other environmental stimuli on biofilm formation.¹⁴¹ Metabolomics can also be used for detecting disease biomarkers,¹⁴² and as a supplementary tool for proteomics and transcriptomics. The linkage between metabolomics, mRNA and protein expression makes it possible to visualize the biological state of an organism.¹⁴³

Metabolomics is the bridge between genotype and phenotype.¹³² Correspondingly, metabolomics provides a better understanding of a disease since it links the pathology to actual changes in the activity of biological processes. Metabolomics provides an approach to diagnose a disease, monitor its progression, evaluate a response to therapy, and identify potential novel drug targets. Thus, metabolomics has a wide range of applications in drug discovery,¹⁴⁴ including toxicology¹⁴⁵ and functional genomics.¹⁴³ The biofilm study can be categorized into either biological science or medical engineering. However, so far, the main focus of our research is on the stage of mechanism exploration. Therefore, metabolomics is mainly used for such purpose. A series of examples addressed below demonstrated how metabolomics can be used to answer fundamental science problems with some extension of future perspectives of how metabolomics can be applied to engineering areas.

NMR metabolomics has been applied to identify biomarkers for cardiac disease,^{146,147} liver disease,¹⁴⁸ respiratory disease,^{149,150} cancer,¹⁵¹⁻¹⁵⁴ and central nervous system disorders,¹⁵⁵⁻¹⁵⁸ among others. Simply, NMR metabolomics provides a means to differentiate between a disease and healthy state or between drug-treated and untreated. Drug discovery or chemical-lead identification is then based on observing the metabolome change from a disease-state to a healthy state or by simply observing that a compound changes the metabolome. For example, Tizianni *et al.* (2011) describes using NMR metabolomics in a high-throughput screening platform (96 well plates) to identify kinase inhibitors.¹⁵⁹ They demonstrate that changes in the lactate/pyruvate ratio in human leukaemia cells (CCRF-CEM) and human ovarian cancer cells (SKOV-3) was successful

in identifying inhibitors to eEF-2, NF-kB, MK2, PKA, PKC and PKG kinases. Similarly, Halouska *et al.* (2012) demonstrate that the *in vivo* mechanism of action of a chemical lead can be inferred by comparing the metabolome changes to a known drug.¹⁴⁰ If two or more drugs have a similar impact on the metabolome then the compounds share a similar target. Additionally, NMR metabolomics is also widely-used for drug development and personalized medicine.¹⁶⁰⁻¹⁶² The consortium for metabonomic toxicology (COMET), an organization of major pharmaceutical companies, was formed to share metabolomics data from drug studies to characterize metabolites associated with kidney or liver toxicity. The overall protocol for toxicity analysis is comparable to drug discovery and lead identification. Specifically, biofluid (urine, serum, saliva, *etc.*) metabolites from animals or patients undergoing drug treatment are analyzed by NMR to identify metabolites known to be associated with drug toxicity or disease biomarkers. If the biofluid metabolites reveal the presence of drug toxicity or the lack of drug efficacy, then an alternative treatment can be prescribed.

Metabolomics is routinely combined with alternative strategies to resolve a biological problem and to generate a comprehensive analysis. Although different strategies may require alternative experiments and data analysis, targeted metabolite analysis, metabolic profiling, and metabolic fingerprinting share the same general workflow from sample preparation to data collection and analysis. Targeted metabolite analysis is the exclusive study of the direct product of a corresponding enzyme or protein.¹³² Metabolic profiling is the identification and quantification of a set of pre-defined metabolites in a biological sample.¹⁶³ The metabolites may belong to a specific

class of compounds or a particular metabolic pathway. As an illustration, metabolic fingerprinting can be used to probe different metabolic phenotypes. Metabolic profiling can then be used to provide a detailed analysis of specific metabolite changes between the two phenotypes. Metabolomics can also be referred to as “metabolic fingerprinting”,¹⁶⁴ and is generally meant to rapidly classify biological samples. The combination of different strategies makes metabolomics a flexible and versatile technique for the analysis of various biological systems, such as bacterial biofilms.^{165,166} The NMR metabolomics methodologies described for drug discovery are equally applicable to investigating bacterial biofilms, for identifying new drug targets and chemical leads, and evolving lead candidates into new drugs. Again, characterizing and comparing the metabolomic differences between planktonic cells and biofilms provides a means to identify active and relevant biological processes associated with biofilm formation. Correspondingly, proteins involved in these pathways are potential drug targets. Identifying and validating drug leads can then be accomplished by observing chemical-induced changes in the metabolome related to biofilm formation.

NMR-based metabolomics generally refers to a comprehensive approach to the analysis of metabolomic samples that include specific NMR experiments, sample preparation protocols, and multivariate statistical analysis.¹⁶⁷⁻¹⁶⁹ NMR spectroscopy is used to characterize the metabolic samples by providing both qualitative and quantitative data.¹³⁴ NMR-based metabolomics of bacterial biofilms consist of the general procedures outlined in figure 2.5: 1) prepare the metabolic samples by culturing the desired bacterial strains under *identical* conditions, the only variable should be the specific environmental

or genetic factor being investigated; 2) prepare the NMR samples by lysing the cells, extracting the metabolites, and removing cell debris; 3) detecting the metabolites through various NMR techniques; and 4) spectral processing, data normalization, statistical analysis, and metabolite identification. The success of metabolomics largely depends on accomplishing each step in a highly controlled and uniform manner. Variation in the NMR metabolomics data should result from relevant biological differences between the samples as opposed to artifacts introduced from sample or data handling. For example, extracting the metabolites from the lysed cells should occur quickly and at low temperatures to avoid changes to the metabolome that result from the process of harvesting the cells. In effect, all potential variables such as: the number of cells, growth phase, culture media, experimental conditions, bacterial strain, time, *etc.* need to remain constant between all bacterial samples.¹⁷⁰ Again, the only difference between the various bacterial cultures should be the specific environmental or genetic factor being investigated. Uniformity is the key to a successful metabolomics experiment. It is impractical to investigate a heterogeneous sample composed of multiple bacterial organisms since there is no mechanism to associate the majority of the commonly observed metabolites to a specific organism. Alternatively, using flow cytometry or other techniques to separate the cells prior to investigating the metabolome does not solve the problem.¹⁷¹ The results would be invalid because the time required and the process of separating the cells would perturb the state of the cells and, correspondingly, the metabolome.

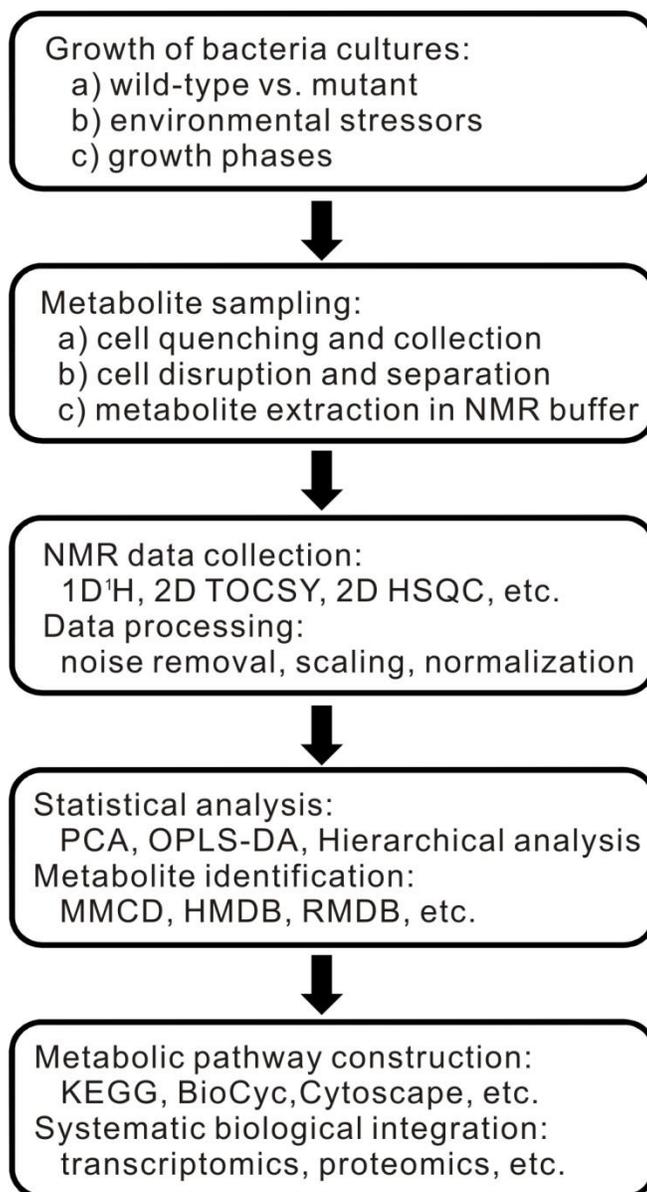


Figure 2.5 A flowchart illustrating the general protocols for an NMR metabolomics study.

Fundamental to a metabolomics study is the identification of the “classes” or “groups” of bacterial cells that will be compared. A straightforward application is the comparison between two groups, a bacterial strain in its planktonic state (class one) and

its biofilm state (class two). Models of higher complexity include even more classes. For example, Figure 2.6 illustrates the application of metabolomics to monitor *in vivo* drug activity. The diagram illustrates the general analysis of clustering patterns in a principle component analysis (PCA) scores plot. Simply, the metabolome of four different classes are compared: 1) wild-type cells, 2) mutant cells where the drug-target has been genetically knocked-out, 3) wild-type cells treated with the drug, and 4) mutant cells treated with the drug. The activity and selectivity of the drug is determined by the relative similarity of the four different metabolomes as described in Figure 2.6. As an illustration, a drug is selective and active if the metabolome of the wild-type cells treated with the drug clusters together with the metabolome obtained from the mutant cells with and without drug treatment, where this cluster forms a separate cluster from the wild-type cells without the drug treatment (Figure 2.6, *top-right*). Simply, these results indicate the protein target in the wild-type cells was chemically inactivated since the metabolome is identical to the mutant cells where the protein was genetically inactivated. It also differs from the wild-type cells without the drug treatment where the protein is still active. The drug is selective because there is no difference between the metabolomes for the mutant cells with or without drug treatment. This analysis can be easily generalized. The “drug” in this scenario can also be taken as any environmental condition, while the “mutant” can be taken as the drug target or any knockout, repressed or over-expressed gene.

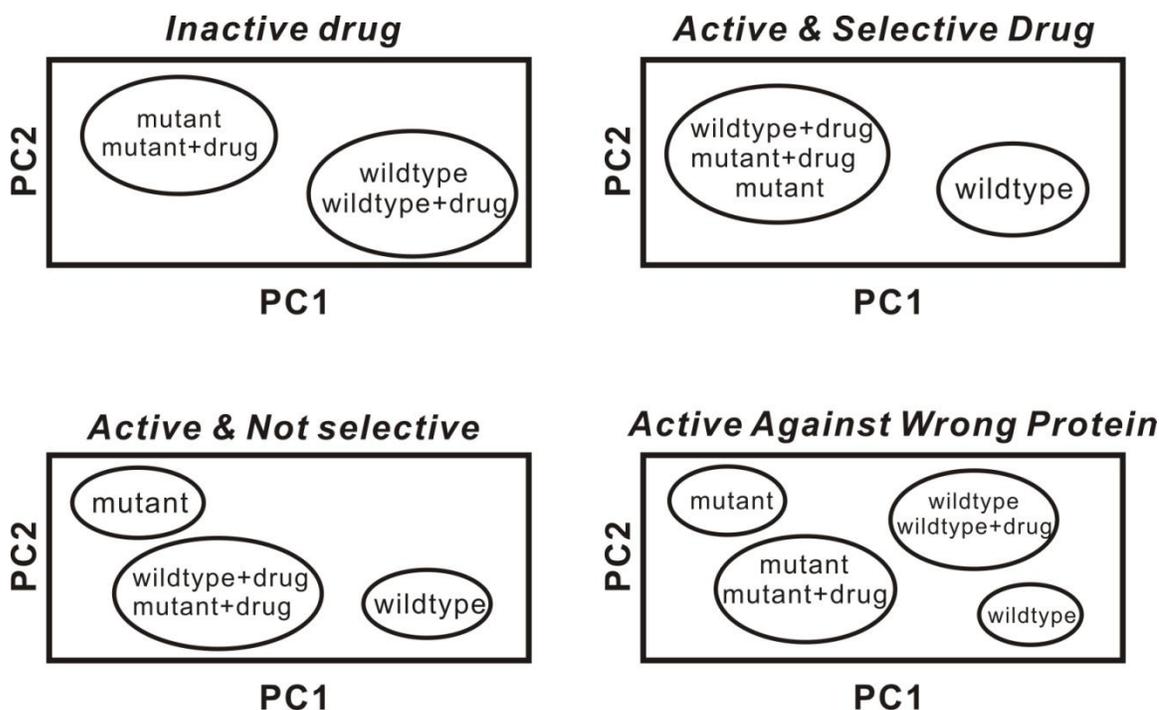


Figure 2.6 Cartoon illustrations of hypothetical PCA scores plot for the following scenarios: inactive compound, active and selective inhibitor, active, nonselective inhibition of target and secondary protein, and active, nonselective preferential inhibition of secondary protein. Reprinted with permission from reference.¹⁷²

2.5 Metabolomics Sample Preparation

Metabolite sample preparation includes cell quenching, cell harvesting, cell disruption and metabolite extraction. An important advantage of NMR-based metabolomics is the minimal and relatively simple sample preparation protocol. Nevertheless, the details of the procedure influence the accuracy, reliability, and reproducibility of the metabolomics data.¹³⁴ Different approaches to sample preparation have various advantages in terms of speed, capability, consistency, efficiency and metabolite recovery yield.¹⁷³⁻¹⁷⁵ Since biofilms can form on a wide range of surfaces or

habitats, the experimental conditions for growing and harvesting cells can be highly variable. Thus, this review will focus on a general discussion of sampling methods for biofilm-related planktonic cells.

A very critical issue in sample preparation is the need to rapidly and efficiently quench all enzymatic and biological activities in order to capture an accurate “snap-shot” of the metabolome. This is because metabolites such as pyruvate, fumarate, oxoglutarate, phosphoenolpyruvate, fructose-6-phosphate, and others have a rapid turnover rate.¹⁷⁶ Also, it is important to avoid inducing a stress response or cell death that would completely invalidate the study. Thus, a quick quenching step that involves reducing the cell temperature has been shown to be a useful approach to slow down enzyme activity within a cell.¹⁷³ Methanol is commonly used because of its low melting point and minimal toxicity relative to other organic solvents.¹⁷⁴ However, methanol may only be suitable for gram-positive bacteria or fungi due to the possibility of cell leakage and the loss of metabolites during the quenching and washing steps.¹⁷⁶ Choosing the proper metabolite extraction protocol is extremely critical since it influences the efficiency and accuracy of the entire metabolomics experiment.

Improperly removing the cell-growth medium and washing the cells is an easy way to contaminate the metabolomic samples and generate unreliable data. Filtration and centrifugation are the two main methods of removing the culture medium before collecting the metabolome. Centrifugation takes longer so there are concerns regarding induced stress and metabolome changes, but it has a higher consistency in sample preparation. Conversely, filtration is significantly faster, and it is also easier to quench

cells on a filter membrane. But, there are also practical concerns with uniformly and consistently retrieving all the frozen cells from the filter paper. Nevertheless filtration-quenching was demonstrated to have the highest yield for an *S. aureus* metabolomics study.¹⁷⁷ Also, an NMR-based metabolomics study of *P. pastoris* applied a single centrifugation step and demonstrated that there was no benefit to including an additional washing step in the quenching process.¹⁷⁸ Directly growing *E. coli* cells on filter paper may provide an efficient and fast approach to quench cells and extract the metabolome.¹⁷⁹ To date, the choice of technique to separate and wash cells is still very flexible, which implies a necessary optimization step for any metabolomics study. An inability to efficiently arrest all cell processes and purify the cells without inducing leakage or lysis will inadvertently lead to undesirable changes in the state of the system and the metabolome. Thus, choosing system appropriate washing, quenching and cell separation protocols is the first and most critical step of a metabolomic project and will determine the validity of the entire study.

Cell lysis and metabolite extraction can sometimes be done simultaneously. Both mechanical disruptions, such as the Fast-Prep system or organic solvent-based methods are widely used.^{175,180} Trichloroacetic acid (TCA) is a traditional approach for lysing cells from filter paper, but TCA causes a significant background for metabolomics data because it also degrades the filter paper.¹⁷⁶ The optimal extraction buffer should: 1) extract the largest number of metabolites, 2) be nonselective and not exclude molecules with particular physical or chemical properties, and 3) be nondestructive and not modify metabolites through chemical or physical means.¹⁸¹ Since metabolites are normally

dissolved in a D₂O buffer or CDCl₃ for NMR, choosing alternative extraction solvents is not particularly beneficial. A common extraction solvent is a (5:2:2) v/v mixture of methanol, chloroform, and water.¹⁸² Others extraction mixtures include a (1:1) v/v mixture of methanol and chloroform, or water and chloroform for metabolite extractions that include lipids.¹⁸³

2.6 NMR Experiments for Metabolomics

There are many practical challenges encountered when studying the bacterial metabolome. A cellular metabolome can contain upwards of thousands of metabolites, with a 7 to 9 order of magnitude range in concentrations (picomoles to millimoles).¹⁸⁴ Therefore, it is generally not possible to analyze all cellular metabolites in a single experiment. Also, cellular metabolism is very sensitive to environmental changes, where the measurement and sampling process can influence the metabolome. Thus, metabolomic measurements are also perturbed by including separation techniques. Correspondingly, each biological system requires experimental optimization to accurately study its metabolome.

Mass spectrometry (MS) and nuclear magnetic resonance (NMR) spectroscopy are the primary analytical techniques used for metabolite detection. Mass spectrometry measures the mass-to-charge ratio of charged molecules that can be used to determine the elemental composition and elucidate the chemical structure of molecules. While an exact mass is determined by MS, the limited molecular-weight dispersion of the metabolome generally requires using chromatography.¹⁸⁵ Gas chromatography, high performance

liquid chromatography (HPLC) and capillary electrophoresis are common separation techniques used by MS-based metabolomics.^{186,187} Unfortunately, the use of chromatography to separate metabolites inevitably leads to changes in the metabolome. The amount that is recovered from the chromatography step will vary for each metabolite, where some metabolites will be lost or chemically modified. MS also requires ionization of the molecule for detection with the corresponding uncertainty that a specific metabolite will ionize. Also, determining a concentration for each metabolite by MS is particularly challenging.

Conversely, NMR determines a molecular structure by measuring nuclear chemical shifts within a magnetic field.¹⁸⁸ Three valuable pieces of information are obtainable from a single peak in an NMR spectrum (Figure 2.7). The chemical shift is related to the local chemical environment of that specific nucleus (^1H , ^{13}C , ^{15}N , *etc.*), and the peak-splitting (J-coupling) identifies chemically bonded nuclei, which are both used to identify the chemical structure of the metabolite.¹⁸⁸ Importantly, the peak intensity is directly proportional to the metabolite's concentration. Typically, multiple distinct NMR resonances are observed per molecule that eliminates the need for chromatographic separation and increases the accuracy of metabolite identification.

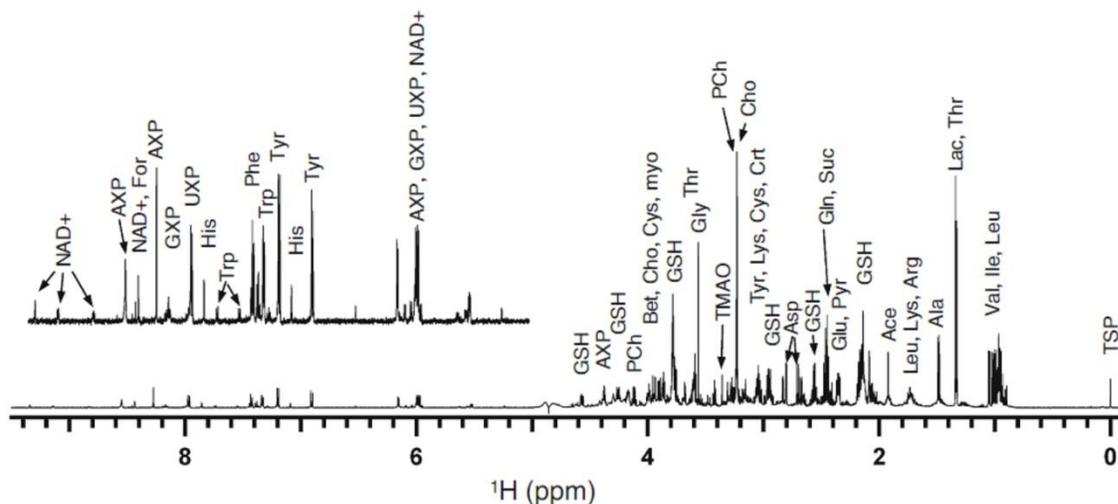


Figure 2.7 ^1H NMR spectra of the intracellular metabolic profiles of aqueous extract of MCF-7 cells. The sample used for spectrum contained $\sim 6 \times 10^5$ cells. The inserted spectra were also plotted using the same expanded x scale. Reprinted with permission from reference.¹⁸⁹

MS is significantly more sensitive than NMR and covers a wider diversity of the metabolome, even though the use of cryogenic probes has significantly increased the sensitivity of NMR by a factor of four.¹⁹⁰ In effect, NMR only detects the most abundant metabolites that are present at concentrations greater than 1 to 5 μM . ^1H NMR is typically used for metabolomics since ^1H NMR is 64 times more sensitive than ^{13}C NMR. Nevertheless, NMR cryoprobes can still be used to detect metabolites using naturally occurring ^{13}C , which has an abundance of only 1.1% (^{12}C is not detectable by NMR).^{191,192} MS is a destructive technique, but it requires a significantly lower sample amount (< 100 L) compared to NMR (600 L). Correspondingly, both approaches are complementary to each other and contribute inherently distinct information to the analysis of a metabolome.¹⁹³ The complementary nature of MS and NMR has been

demonstrated by a number of metabolomic studies using both techniques.¹⁹⁴⁻¹⁹⁸ In effect, the MS and the NMR data can be combined to create a three-dimensional scores plot. The added dimensionality from complementary data may provide the additional resolution necessary to differentiate between multiple classes or groups.

The application of NMR spectroscopy for metabolomics can be categorized into one of three groups, one dimensional (1D) NMR, two dimensional (2D) NMR experiments and solid state NMR.¹⁸⁸ 1D and 2D solution-state ^1H NMR experiments are commonly used for global metabolomics analysis of bacterial cell extracts.¹⁹⁹ Conversely, solid state NMR can be used to analyze intact cells.²⁰⁰⁻²⁰² In addition to ^1H , other nuclei are also used in 1D NMR-based metabolomics, such as the metabolic profiling of the carbohydrate cycle using 1D ^{13}C NMR,^{203,204} or tissue metabolism using 1D ^{31}P NMR.²⁰⁵

A typical 1D ^1H NMR spectrum of a bacterial cell lysate may contain thousands of sharp lines from low molecular weight metabolites (Figure 2.7).¹³⁴ The entire 1D ^1H NMR spectrum is used as a “fingerprint” to characterize the state of the bacterial cell. A global investigation of the metabolome is simply based on a comparative analysis of the features present or absent in each 1D ^1H NMR spectrum. Effectively, a global metabolomic analysis is based on how similar or how different the 1D ^1H NMR spectra are between each class or group. It is not necessary to assign each 1D ^1H NMR spectrum to identify and quantify all the metabolites present in each sample. Overlapping these relevant NMR resonances and interfering with the analysis may be broad bands from proteins or other biomolecules, which can be readily removed by using a Carr–Purcell–Meiboom–Gill (CPMG) spin-echo sequence.²⁰⁶ The CPMG pulse sequence takes

advantage of the large difference in T_2 relaxation times between small molecular-weight metabolites and large MW biomolecules. Effectively, the NMR resonances from the biomolecules rapidly decay during the CPMG pulse. Alternatively, proteins and other biomolecules can be removed by an appropriate choice of extraction solvents.^{207,208} The large interfering signal from water or other buffer components is also eliminated by the use of appropriate NMR solvent suppression methods and 100% deuterated buffer.²⁰⁹ Basically, the water NMR resonance is set in the center of the spectrum where selective irradiation and gradient pulses suppress the solvent peak while leaving all other peaks unaffected. In effect, any resonance in the 1D ^1H NMR spectrum that does not originate from the bacterial metabolome will generate a “false feature” that needs to be removed. NMR resonances originating from proteins or solvents are likely to be variable and dominate the spectra relative to metabolite signals. This will lead to an erroneous interpretation of the 1D ^1H NMR spectra and incorrect sample classification. For instance, replicate samples may not cluster together because of a significant variation in the peak height and peak shape of the water resonance despite the overall similarity in all the metabolite NMR peaks.

High resolution magic angular spinning (HR-MAS) is used to generate *in vivo* 1D ^1H spectra of solids comparable to solution-state NMR.²¹⁰ Thus, small, intact, and untreated cells or tissues can be directly analyzed by HR-MAS by spinning samples at speeds between 4 and 12 kHz at the “magic” angle of 54.7° relative to the external magnetic field. Spinning the sample significantly reduces NMR line widths by averaging out chemical shift anisotropy, magnetic susceptibility, and dipolar coupling that are

prominent in solid samples.¹⁸⁸

Assigning a 1D ^1H NMR spectrum to identify the metabolites present in a metabolomics sample is challenging because of the large number of peaks, the significant overlap in peaks, the high chemical shift degeneracy (multiple metabolites have some chemical shifts in common), and an incomplete database of NMR reference spectra for metabolites. Again, assigning a 1D ^1H NMR spectrum is not necessary for a global analysis of the metabolome, but identifying the specific metabolites that are changing and are the main contributors to class distinction is extremely valuable for understanding the underlying biological differences. Statistical total correlation spectroscopy (STOCSY) can be used to associate multiple NMR peaks from the same molecule in a complex mixture.²¹¹ This significantly simplifies the assignment problem since most, if not all, of the NMR resonances for a given metabolite can be used together in a database search. A positive identification only occurs when all of the observed chemical shifts match the metabolite's known chemical shifts in a database. In STOCSY, a series of 1D ^1H NMR spectra is converted into a pseudo-2D spectrum that is based on a correlation of peak intensities. NMR peaks from the same metabolite will change together as the metabolite's concentration varies across multiple distinct classes. The statistical heterospectroscopy (SHY) is similar in concept to STOCSY.²¹² Instead of correlating NMR peak intensities, SHY correlates chemical shifts from NMR with mass-charge (m/z) data from mass spectroscopy. Thus, SHY can improve molecular identification by directly cross-correlating NMR chemical shifts with molecular weight.

More commonly, 2D NMR spectroscopy improves the accuracy of metabolite

assignments by significantly increasing spectral resolution by extending chemical shift information into a second frequency dimension. Additionally, 2D NMR experiments can identify the network of resonances associated with a specific metabolite through J-coupling. Two-dimensional correlation spectroscopy (COSY) and total correlation spectroscopy (TOCSY) experiments identify spin–spin coupling connectivities that identify chemically bonded pairs of hydrogens, carbons or nitrogens.^{213,214} To a lesser extent, 2D J-resolved NMR experiments are also used for metabolomics studies.²¹⁵ In a 2D J-resolved NMR experiment, the data is dispersed into two-dimensions based on chemical shifts and the J-coupling pattern.²¹⁶ While it is possible to match coupling patterns to identify bonded resonances, this is generally not practical for complex metabolomics data set. So, a 2D J-resolved NMR experiment has significantly less information than a 2D COSY or TOCSY experiment while requiring the same amount of experimental time. Alternatively, a 1D projection of the 2D J-resolved NMR experiment can be used to simplify the NMR spectra by removing peak multiplicity due to J-coupling.²¹⁷ This dramatically reduces the number of peaks and correspondingly reduces peak overlap. Removing peak splitting improves the accuracy of metabolite identification and quantification. Of course, a single 2D NMR experiment may require an hour or longer to acquire compared to a few minutes for a 1D NMR experiment.

The 2D ^1H , ^{13}C HSQC (Heteronuclear Single-Quantum Correlation) experiment correlates the ^1H and ^{13}C chemical shifts for each C-H pair in a molecule.²¹⁸ This provides unique information relative to the COSY, TOCSY or J-resolved experiments. Also, metabolite assignments are easier with a 2D ^1H , ^{13}C HSQC experiment because of

two correlated and distinct chemical shift assignments, the large ^{13}C chemical shift dispersion, and the simplified spectrum without splitting from J-coupling. However, due to the low natural abundance (1.1%) of ^{13}C , ^{13}C -labeled compounds, such as ^{13}C -methanol, ^{13}C - CO_2 ²¹⁹, ^{13}C -glycerol²²⁰ and ^{13}C -glucose²²¹ are required as a bacterial carbon source to enhanced the sensitivity of the NMR spectrum.²²² This significantly simplifies and focuses the analysis of the metabolome. Only metabolic intermediates and products of the ^{13}C -labeled materials will be partially or completely enriched with ^{13}C . Correspondingly, only these metabolites will be observed in a 2D ^1H , ^{13}C HSQC spectrum, which provides a means to follow carbon-flow through the metabolome and identify the perturbed metabolic pathways. Standard HSQC experiments are not quantitative because of significant variability in coupling-constant and relaxation times (T_1 and T_2) between metabolites. Nevertheless, the newly developed 2D extrapolated time-zero ^1H , ^{13}C HSQC (HSQC₀) experiment allows for the calculation of metabolite concentrations.²²³ Basically, a series of 2D HSQC spectra are collected with an increasing number of the core HSQC experiment. The peak intensity will decrease linearly with the number of HSQC blocks, where a linear fit and extrapolation back to zero HSQC blocks will determine the true peak intensity and metabolite concentration. Examples of 2D NMR spectra used to identify metabolites from biological samples are shown in Figure 2.8.

Figure 2.8 a) Overlay of 2D ^1H , ^{13}C HSQC spectra comparing wild-type *S. epidermidis* strain 1457 (red) and aconitase mutant strain 1457-*acnA::tetM* (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 ^1H , ^1H TOCSY spectra comparing wild-type *S. epidermidis* strain 1457 (red) and aconitase mutant strain 1457-*acnA::tetM* (black) grown for 6 h in standard TSB media. Reprinted with permission from reference.¹⁴¹

For chemometrics (please see the *Chemometrics and Bioinformatics Analysis of Metabolomics Data* section), the 1D ^1H NMR spectra are transformed into a data matrix of integrated peak intensities and corresponding chemical shift values. This data is used to define the classes and to identify NMR spectral features that differentiate the classes. Unfortunately, subtle instrument, temperature, and sample condition variability (pH, ionic strength, *etc.*) can result in chemical shift differences between replicate samples. Correspondingly, misalignments will occur in the NMR data matrix between these replicate samples that will lead to clustering errors independent of any biological differences. One approach to normalize NMR metabolomics samples is the inclusion of a known concentration of chemical shift reference compound such as the sodium salt of 3-trimethylsilylpropionic acid (TSP). The TSP peak intensity can also be used to calibrate the concentrations of the metabolites in the biological sample. Unfortunately, an internal reference only corrects for global differences. It does not correct for individual peak position and shape variability due to subtle experimental differences. Instead, binning or bucketing is commonly used to correct for peak variability between replicate samples.^{224,225}

In its simplest implementation, bucketing divides the NMR spectrum into equal sized, non-overlapping sub-spectral regions with a width or bin size of 0.025 ppm. So, a typical 1D ^1H NMR spectrum with a width that ranges from 0 to 10 ppm will have 400 buckets. Bin sizes typically range from 0.01 to 0.04 ppm. The peak intensity within each bucket is integrated, where these resulting integrals are then used as input to the chemometrics analysis. In this manner, the buckets or bins smooth-out small peak

variability with the expectation that the same peaks occur in the same bucket.

Unfortunately, bin edges create a second problem; the undesirable splitting of a peak between buckets. Thus, “intelligent” or “adaptive” bucketing techniques have been developed that vary the individual bin size to avoid dividing peaks between multiple buckets.²²⁶⁻²²⁹ These methods use a Gaussian function, a recursive algorithm, optimize an objective function using a dynamic programming strategy, or use underestimated wavelet transforms to automatically identify bin edges. In all cases, intelligent bucketing performs significantly better than uniform bucketing, where dynamic adaptive binning was recently shown to perform the best.²²⁹ The use of intelligent bucketing results in a significant improvement in replicate clustering in scores plots since it minimizes spectral differences that are biologically irrelevant. Correspondingly, noise regions of the NMR spectra are typically zeroed or removed from the bucketing.²³⁰ Similarly, buckets resulting from solvent or buffer peaks that are unrelated to the bacterial metabolome are also excluded. Again, this eliminates class distinction that results from biologically irrelevant data. In essence, the variables used in the chemometrics analysis should be relevant to real variations in the metabolome between the classes.¹⁷⁰

Peak alignment is a more robust and complex alternative to NMR binning.²³¹⁻²³⁴ The goal is to remove the chemical shift variability between the replicate 1D ¹H NMR spectra by individually aligning each NMR peak to a representative spectrum from each class. The approaches used include fuzzy warping, genetic algorithms, a generalized fuzzy Hough transform approach, a reduced set mapping (PARS) algorithm, or a recursive segment-wise peak alignment. Each approach demonstrated acceptable results

on test metabolomic samples and was shown to improve upon the results obtained from bucketing.²³⁴ Nevertheless, intelligent bucketing is still the predominate methodology used in NMR metabolomics.

An NMR spectrum is experimentally collected as a free induction decay (FID) that requires further processing to convert the time-domain data into a frequency-domain spectrum through a Fourier transform. This process can be accomplished using a variety of software packages (Table 2.1). All of the software packages can import popular NMR data formats such as: Bruker, Varian, Jeol Delta, JCAMP-DX, and others. Some NMR software can process both 1D and 2D data, where others only focus on processing 1D (most popular) or 2D data sets. In addition to the Fourier transform step, the processing of NMR data may include zero-filling, phase correction, baseline correction, applying a window function, and removal of solvent peaks. The inclusion of any of these steps induces changes in the resulting NMR spectrum that is not biologically relevant. Thus, uniformity in the NMR processing protocol is essential, where minimizing all spectral manipulation is ideal. In general, phase correction is essential in order to obtain purely absorption peak shapes. Uniform zero-filling of the NMR spectra is typically acceptable since it provides a constant improvement in the digital resolution. Similarly, removing residual solvent peaks by simply zeroing the corresponding region of the NMR spectrum does not have any detrimental effect because these solvent regions are not included in the binning process. Conversely, applying a window function or incorporating baseline correction should be avoided, if possible, since these processes induce significant variable changes in the NMR spectrum. The goal of a window function is to either

increase the spectral resolution or the signal-to-noise by multiplying the FID with a mathematical function. As a result, each NMR peak shape is artificially changed. Instead, the signal-to-noise and spectral resolution should be improved experimentally. A baseline correction is required if the baseline for the NMR spectrum is not flat. A flat baseline is necessary for reliable chemometrics analysis. Distorted baselines may result from intense solvent or buffer peaks, from phasing problems, errors in the first data points of the FID, large range of T_1 values (short recycle times), or short acquisition times. Again, baseline problems should be experimentally minimized. Nevertheless, numerous computational approaches to obtain a flat baseline are available and include a linear or polynomial (up to 6 orders) fit of the baseline, FID reconstruction and spectrum averaging. The choice of a baseline correction method depends on the specifics of the baseline distortion, which is typically determined by trial-and-error. Importantly, a single baseline correction method must be used for an entire metabolomics dataset in order to avoid inducing class distinctions that are primarily a result of the NMR processing protocol.

Table 2.1 NMR software packages.

Software	OS	Source	Reference
ACDLab	Win	Commercial	1D, 2D Advanced Chemistry Development, Inc.
CCPNmr	Win/Mac/Unix	Academic	2D ²³⁵
Felix	Win/Unix	Commercial	2D Felix NMR, Inc.
FERCH	Win	Commercial	1D PERCH Solutions Ltd.
Gifa	Unix	Academic	1D, 2D ²³⁶
iNMR	Mac	Commercial	1D, 2D Nucleomatica, Inc.
matNMR	Win/Mac/Unix	Academic	1D, 2D ²³⁷
MestreNova	Win/Mac/Unix	Commercial	1D, 2D Mestrelab research
NMRPipe	Win/Mac/Unix	Academic	2D ²³⁸
NMRViewJ	Win/Mac/Unix	Commercial	2D One Moon Scientific, Inc.
NUTS	Win	Commercial	1D, 2D Acorn NMR Inc.
RMN	Mac	Academic	1D, 2D P. J. Grandinetti, Ohio State University, Columbus, OH, U.S.A.
rNMR	Win/Mac/Unix	Academic	2D ²³⁹
Sparky	Win/Mac/Unix	Academic	2D T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco, CA, U.S.A.
SpinWorks	Win/Mac/Unix	Academic	1D, 2D K.Marar, University of Manitoba, Winnipeg, Manitoba, Canada
TARQUIN	Win/Mac/Unix	Academic	Solid state ²⁴⁰
WIN-NMR	Win	Commercial	1D Bruker Instruments, Inc.

Ideally, the overall concentration of the metabolites and the corresponding signal-to-noise of each replicate 1D ¹H NMR spectrum will be essentially identical.

Unfortunately, in practice, there may be a significant variability in the signal-to-noise between replicate NMR spectra due to random errors in cell lysing, metabolite extraction, or the number of bacterial cells per sample. Correspondingly, the 1D ¹H NMR spectra needs to be normalized.²⁴¹ Center averaging is a common normalization technique:

$$Z = \frac{X_i - \bar{X}}{\sigma} \quad (1)$$

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. Other normalization techniques include: normalization to a constant sum, Pareto scaling, mean-centering, logarithmic scaling, probabilistic quotient normalization, *etc.* A recent analysis of NMR processing procedures by De Meyer *et al.* (2010), suggests that a combination of intelligent bucketing with probabilistic quotient normalization yields the best results.²⁴² Of course, it was not practical to explore all the possible combinations of the various processing techniques and, as a result, the analysis did not include center averaging. This highlights a particular challenge with processing NMR metabolomics data; there is no clear consensus of an optimal protocol because of all the possible combinations of parameters that need to be evaluated.

2.7 Chemometrics and Bioinformatics Analysis of Metabolomics Data

For a detailed understanding of a complex biological system, it is essential to follow the response of an organism to a conditional perturbation at the transcriptome, proteome or metabolome level.¹⁴³ Metabolic products are dependent on genotype, environment, time, and location.²⁴³ Perturbations in any of these factors may lead to a variety of biological changes that inadvertently effect the metabolome.¹⁴³ The primary goal of chemometrics is to reduce the complexity of the NMR-based metabolomics data to understand these global correlations. In essence, chemometrics identifies the major features within an NMR spectrum, the presence or absence of peaks, the change in peak intensity, shape, or the change in chemical shifts that *differ* between the various classes. A class definition can be based on any variable that affects or changes the bacterial cell

culture, such as different bacterial strains (including wild-type and mutant cells) and different experimental conditions (growth phase, drug dosage, media composition, pH, *etc.*). Bioinformatics is used to uncover and characterize all the associated variables and to reveal the underlying relationships. Essentially, bioinformatics is used to identify the metabolites that correlate with the major changes in the NMR spectra and to associate these metabolites with specific proteins and enzymes, with metabolic pathways, and other biological processes.

For statistical modeling, principle component analysis (PCA) is the most common multivariate technique for the comparison of metabolomic data.²⁴⁴ Basically, each multivariable (chemical shift and intensity) NMR spectrum is converted into a single-point in multidimensional Cartesian space (Figure 2.9). Each axis ($V_1, V_2, V_3 \dots V_n$) corresponds to a specific chemical shift where the peak intensity is the value along each axis. PCA determines the vector (\overline{PC}_1) corresponding to the largest variation in the data and fits each NMR spectra to this vector. Each NMR spectrum is assigned a value (PC1-principal component one) based on its distance to the vector (\overline{PC}_1). The process is repeated for the next largest variation (\overline{PC}_2) orthogonal to the first vector. Plotting the resulting principal components (PC1, PC2) corresponds to a scores plot, where similar spectra (and metabolomes) cluster together in a scores plot. Spectra (and metabolomes) obtained from different classes are expected to form separate clusters in the scores plot. Figure 2.9 illustrates the hypothetical separation of class 1 (♦) from class 2 (●) and the relatively tight clustering of the class replicates. The class separation is based on differences in the 1D ^1H NMR spectra.

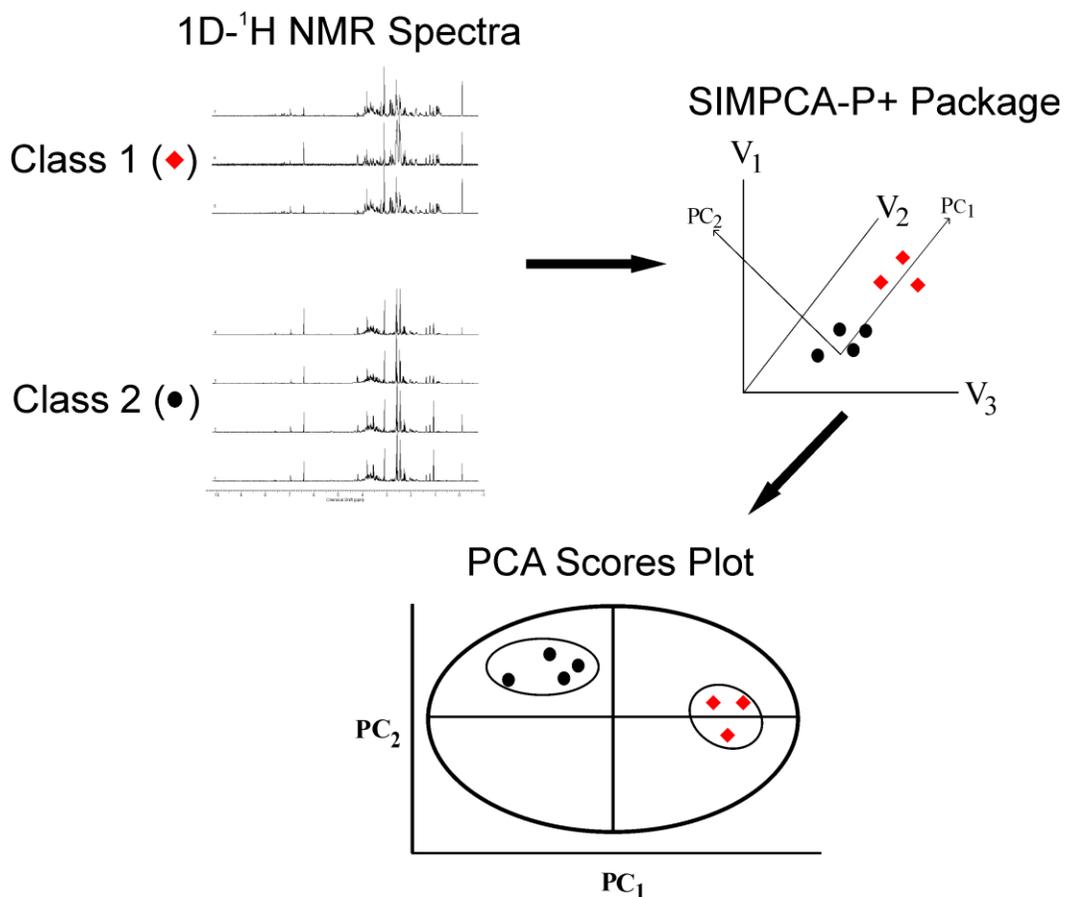


Figure 2.9 Conceptual illustration of the principal component analysis of NMR spectral data. Each 1D ¹H NMR spectrum is converted into a single point in multidimensional space. Each axis (V_1 , V_2 , V_3 , *etc.*) corresponds to a chemical shift (ppm) and the value along the axis is the peak intensity or bin integral. A vector (PC_1) corresponds to the largest variation in the data. The second vector (PC_2) is orthogonal to the first vector and corresponds to the second largest variation in the data. The scores corresponds to the fit of each point (spectrum) to each principal component vector, where the resulting 2D scores plot identifies the relative similarities and differences between the NMR spectra based on the clustering pattern.

Orthogonal partial least squares discriminant analysis (OPLS-DA) is a related, but alternative approach to PCA. Unlike PCA, OPLS-DA is a supervised method. This is an

important and significant difference. In PCA, the different classes are not identified and each 1D ^1H NMR spectrum follows the protocol described above. Thus, the clustering of each NMR spectrum in the scores plot is *solely* determined by the intrinsic differences and similarities in the features of its NMR spectrum relative to all the other NMR spectra. Conversely, in OPLS-DA each 1D ^1H NMR spectrum is assigned to one of two classes, (biofilm vs. planktonic, healthy vs. diseased, treated vs. untreated, wild-type vs. mutant, *etc.*). More than two class definitions can be used, but it is generally not recommended. As a result, OPLS-DA determines the clustering patterns in the scores plot based on the NMR spectral features *and* the class definition. OPLS-DA tries to maximize the separation between the classes based on these manual class assignments, while minimizing the within class variation. This results in tighter class clustering in the OPLS-DA scores plot relative to PCA. Fundamentally, if two classes are defined in OPLS-DA then two clusters corresponding to the two assigned classes will be generated in the resulting scores plot. OPLS-DA assesses a relationship between an X matrix (NMR data) and Y matrix (the 0 or 1 class designation). Thus, OPLS-DA will only identify all the spectral features that can be used to distinguish the two classes regardless of any real significance (i.e., noise, small random variability in peak height, *etc.*). Conversely, the class separation in a PCA scores plot depends on a combination of all principal component variables. Since OPLS-DA is a supervised method or biased by the class assignment, it is essential that the model is validated.¹⁷⁰ Is the class separation due to relevant changes in the 1D ^1H NMR spectra resulting from changes in metabolite concentrations?

The quality of the OPLS-DA model can be evaluated by multiple statistical factors and cross-validation, the goodness of fit (R^2) and the quality assessment score (Q^2).²⁴⁵ A good OPLS-DA model has R^2 values ≥ 0.5 (range of 0 to 1) that is conceptually similar to simple linear regressions. Similarly, a typical value for Q^2 for a biological model is ≥ 0.4 . The leave-one-out cross-correlation technique is commonly used to calculate Q^2 , where a sub-set of the NMR spectra are left out to calculate a model that is then used to predict the left out data.^{246,247} Q^2 is the consistency between the predicted and original data. Importantly, Q^2 and R^2 values should only be used as a figure of merit and not validation of the model. A permutation tests is one approach to validate the model.^{246,247} Simply, the NMR spectra classifications are randomly assigned creating, in principal, a random data set that should have poor class separation and low Q^2 scores. The process is repeated numerous times (> 1000 permutations) until a Gaussian distribution of Q^2 scores are obtained for the random data set. The statistical significance of the model with the correct NMR spectra classification can then be obtained by comparing the model Q^2 with the random Q^2 scores and calculating a P-value. In addition to Q^2 , a similar validation can be achieved by using the area under a receiver operating characteristic (ROC) curve or the number of misclassifications.

Further validation of the OPLS-DA model can be assessed by determining if the NMR features that determine the class separation are biologically relevant. S-plots and loading plots are additional outcomes of an OPLS-DA model. An S-plot identifies the relative contribution of each bin (chemical shift, metabolite) to the clustering in the corresponding scores plot, each bin with a covariance of greater than 0.10 or less than -

0.10 are identified as major contributors to the class separation.²⁴⁸ Similarly, a loadings plot displays the relative contribution of each bin to the principal components. Thus, a valid OPLS-DA model is supported by being able to assign metabolites to the bins that are associated with the class separation. The model is further validated if multiple bins assigned to the same metabolite are all major contributors to the class separation.

Additionally, the identification of multiple correlated metabolites, members of the same metabolic pathway for instance, that are all major contributors to the class separation would further strengthen the reliability of the OPLS-DA model. Again, the overall goal of NMR-based metabolomics is to identify major changes in the NMR spectra that can be associated to a specific set of metabolites with a relationship to a biological process, such as bacterial biofilm formation.

A number of metabolomics software packages have been developed to automate the chemometrics analysis by combining data normalization, data reduction, model prediction and validation, and even metabolite identification, into a single work-flow. These programs include: Automics,²⁴⁹ HiRes,²⁵⁰ MetaboAnalyst,²⁵¹ and the R-package Metabonomic.²⁵² Among these, Automics is the most versatile and extensive package and includes nine different statistical methods applicable to metabolomics data: feature selection (Fisher's criterion), data reduction (PCA, LDA, ULDA), unsupervised clustering (K-Mean) and supervised regression and classification (PLS/PLS-DA, KNN, SIMCA, SVM). Automics also incorporates processing tools to generate a STOCYSY spectrum.

Interpreting a PCA or OPLS-DA scores plot is fundamentally a cluster analysis. If two or more classes cluster together in the scores plot, then the NMR spectra and the corresponding metabolome are considered statistically similar. Conversely, two or more classes that form distinct clusters indicate the samples are significantly different metabolomes. For a simple metabolomics study that involves only two or three different types of samples, the cluster analysis is generally straightforward. An ellipse that corresponds to the 95% confidence limits from a normal distribution for each cluster can be used to define each class in the 2D scores plot to visually determine class separation. Alternatively, Werth *et al.* (2010) recently demonstrated the application of metabolomic tree diagrams combined with standard boot-strapping techniques as a more robust statistical analysis of clustering patterns in scores plots.^{253,254} Simply, each PC value (PC1, PC2) is treated as an axis in a Cartesian coordinate system. An average position is calculated for each class cluster, which is then used to calculate a Euclidean distance between each class to create a distance matrix. The cluster centers and distances between clusters are re-calculated by randomly selecting data points from each class. The process is repeated until 100 different distance matrices are generated, which are then used to generate 100 tree diagrams using Phylip 3.68b.²⁵⁵ A consensus tree is created, where the bootstrap number is simply the number of times each node appears in all 100 trees. Bootstrap values below 50% imply a statistically insignificant separation. Further analysis to identify the metabolic processes that led to class separation may require bioinformatics.²⁵⁶ An Example 2D OPLS-DA scores plot and associated metabolomics tree diagram is shown in Figure 2.10.

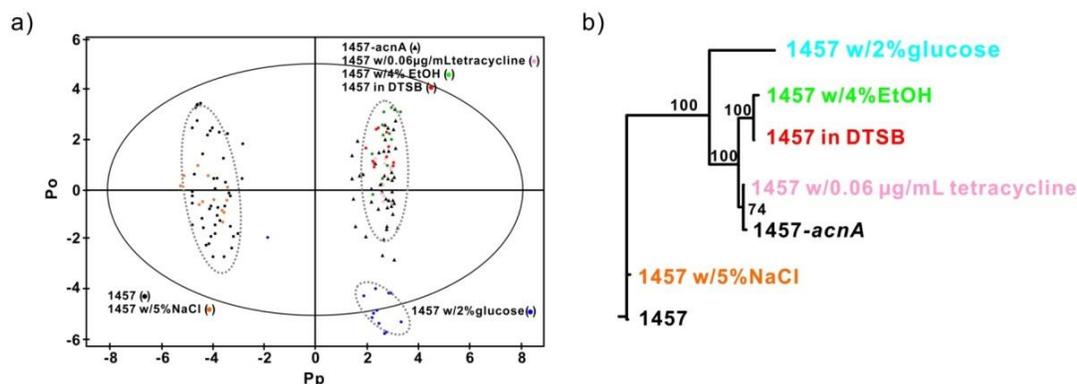


Figure 2.10 2D OPLS-DA comparing wild-type *S. epidermidis* 1457 cells grown 6 h in standard TSB media (●), with *S. epidermidis* 1457 cells grown 6 h in iron-depleted media (DTSB) (●), with the addition of 4% ethanol (●), with the addition of 2% glucose (●), with the addition of 0.06 µg/mL tetracycline (●), with the addition of 5% NaCl (●), and 6 h growth of aconitase mutant strain 1457-*acnA*::*tetM* 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. (b) Metabolomic tree diagram generated from the the 2D OPLS-DA scores plot depicted in (a). 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. Reprinted with permission from reference.¹⁴¹

Metabolic pathway reconstructions has been widely employed with five major goals: 1) contextualization of high-throughput data, 2) guidance of metabolic engineering, 3) directing hypothesis-driven discovery, 4) interrogation of multi-species relationships, and 5) network property discovery.²⁵⁷ NMR-based metabolomics are routinely used to

generate these metabolic networks for a particular biological system (Figure 2.11). The first step of the process is to deconvolute and decode the NMR spectra. This is a difficult and time-consuming process because the high-degeneracy of NMR chemical shifts makes unambiguously assigning an NMR resonance to a specific metabolite extremely challenging. This is further complicated by the incompleteness of metabolomics data and by several technical problems that include proper peak assignment,¹⁹⁹ peak alignment,^{258,259} absolute concentration determination,^{223,260} and intensity normalization.²⁶¹ A number of metabolomics databases that contain NMR and MS spectra with the associated assignments are available: Metabominer,²⁶² Madison Metabolomics Consortium Database (MMCD),²⁶³ BioMagResBank,²⁶⁴ and Human Metabolome Database.²⁶⁵ Some efforts to automate the process have also been made.^{199,249}

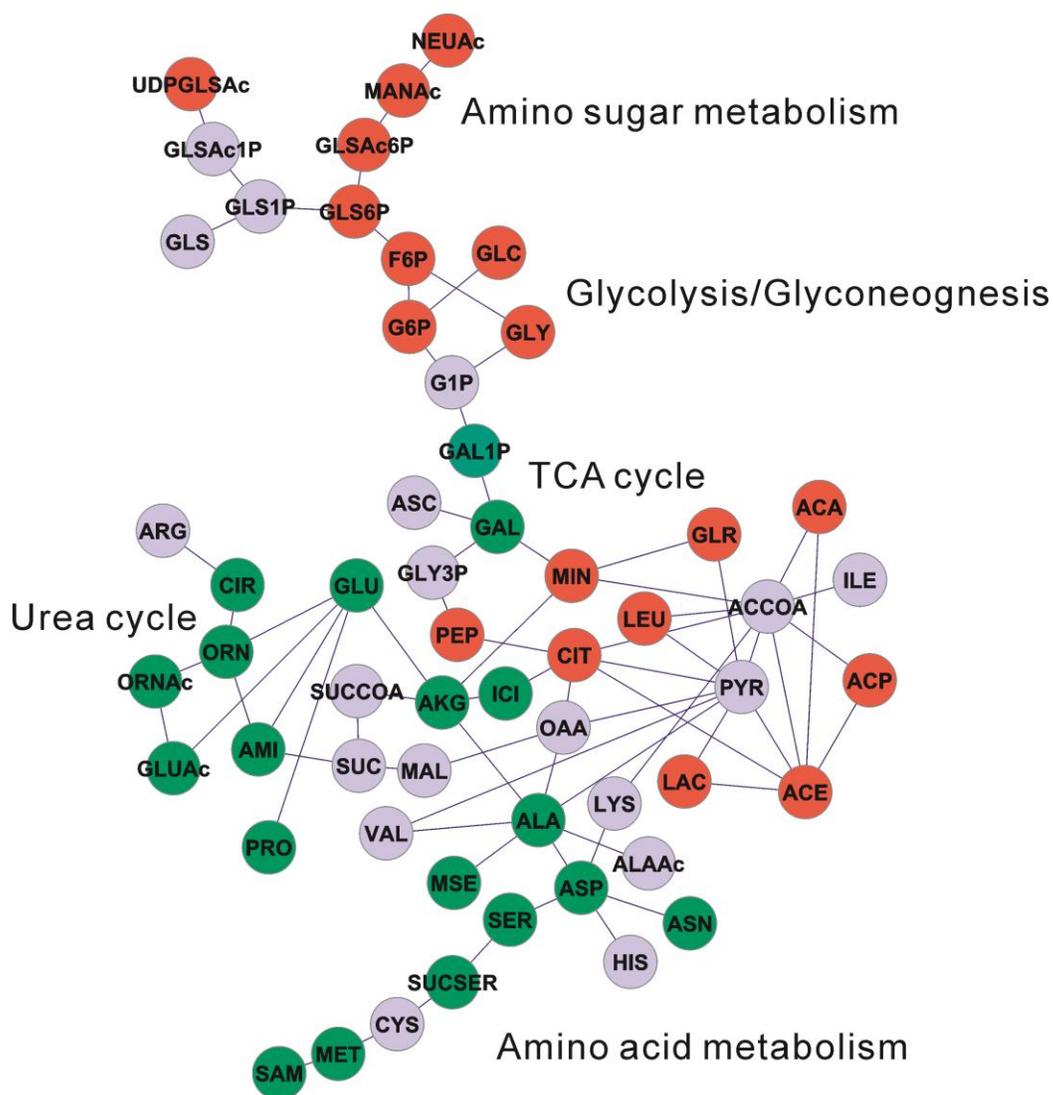


Figure 2.11 Cytoscape network depicting the metabolite concentration changes caused by the inactivation of the TCA cycle. Nodes colored red correspond to metabolites with an increase in concentration due to TCA inactivated. Nodes colored green correspond to metabolites with a decrease in concentration 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 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 (R-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-neuraminic acid), 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-glucuronate), UDPGLSAc (UDP-NAc-D-glucosamine), VAL (Valine). Reprinted with permission from reference.¹⁴¹

The next critical step is to analyze the identified metabolites and find all possible correlations. The presence of metabolites and metabolic pathways in a particular organism can be verified by the KEGG²⁶⁶ and MetaCyc databases.²⁶⁷ A metabolic network map can then be generated using Cytoscape.²⁶⁸ Also, there are many metabolome simulators that are useful for predicting the networks involving hundreds of metabolites.²⁶⁹ Additionally, Cell Designer is a valuable diagram editor for drawing biochemical networks.²⁷⁰

2.8 Analysis of Biofilms with NMR-based Metabolomics

An obvious application of NMR-based metabolomics has been used to characterize the difference in the metabolome between planktonic and biofilm cells.

Gjersing *et al.* (2007) applied ^1H HRMAS to study the metabolome difference between *Pseudomonas aeruginosa* chemostat planktonic and biofilm cells.¹⁶⁶ There was no apparent difference in the extracellular metabolite composition when planktonic and biofilm cells were grown under continually feed chemostat mode. Conversely, the 2D PCA scores plot of the intracellular metabolome indicated a clear distinction between the batch and chemostat planktonic and biofilm cells (Figure 2.12 (*top*)). The corresponding loadings plot suggests major and complex differences between the two metabolomes (Figure 2.12 (*bottom*)). The metabolite differences were not analyzed in detail, but it was noted that biofilm metabolites were generally lower in concentration, possibly due to the cells closer to the substrate operating at a lower metabolic rate. Similarly, Workentine *et al.* (2010) used ^1H NMR to characterize the phenotype of different biofilm colonies of *Pseudomonas fluorescens*.²⁷¹ PCA of the 1D ^1H NMR spectra yielded a 3D PCA scores plot with distinct clustering for the four different *P. fluorescens* colonies (Figure 2.13 (*left*)). The class distinction was attributed to changes in amino acids (Asp, Glu, Gly, Met, Phe, Pro, Trp, Val) and central metabolites (Acetate, Glutathione, Pyruvate). This is suggestive of changes in exopolysaccharide production, response to oxidative stress, and an impaired amino acid metabolism. Interestingly, the four different *P. fluorescens* phenotypes exhibited distinctly different metal sensitivity (Figure 2.13 (*right*)). Unfortunately, it was not possible to determine if the metabolome differences were a result of the different metal sensitivity. Booth *et al.* (2011) also analyzed *P. fluorescens* cells under metal stress, but compared planktonic to biofilms cells.¹⁶⁵ 1D ^1H NMR was combined with GC-MS to obtain a detailed analysis of metabolite changes caused by the

addition of copper (Figure 2.14). Planktonic cells responded differently to copper stress compared to biofilms. Planktonic cells experienced an oxidative stress response as indicated by changes in the TCA cycle, glycolysis, pyruvate and nicotinate and niacotinamide metabolism. Conversely, biofilms exhibited changes in exopolysaccharide related metabolism suggestive of a protective response instead of the reactive changes that occurred in planktonic cells.

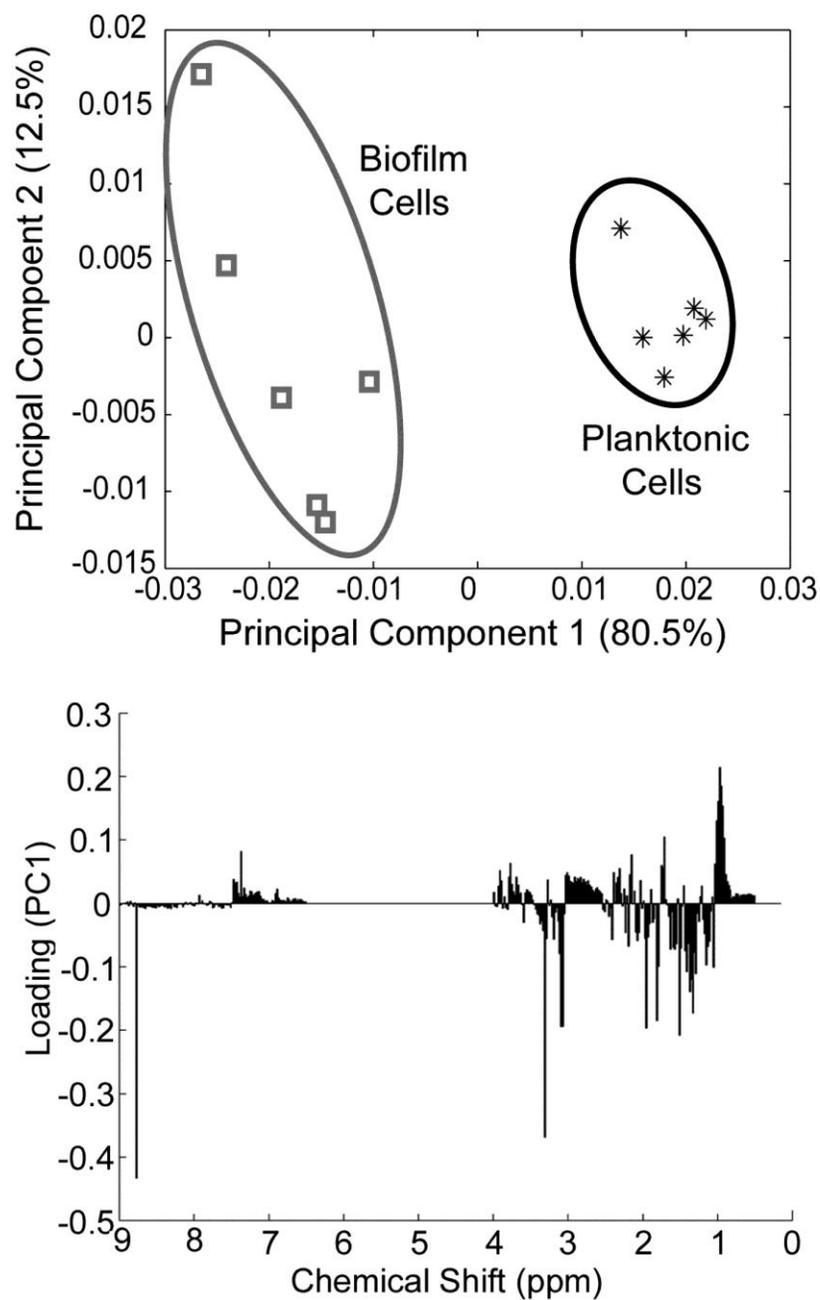


Figure 2.12 (top) PCA score plot for chemostat planktonic and biofilm cell samples for *Pseudomonas aeruginosa*. (bottom) Loading plot of the first principal component. Reprinted with permission from reference.¹⁶⁶

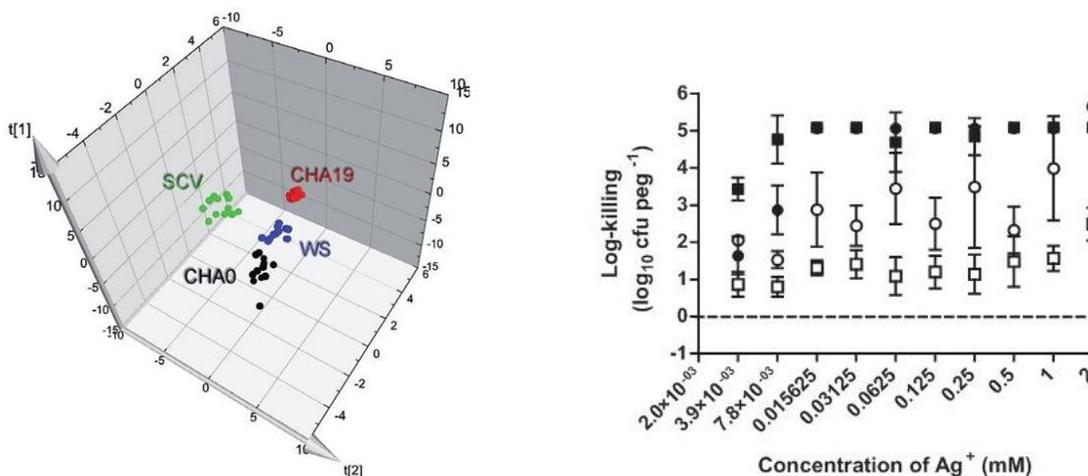


Figure 2.13 (above) Partial least squares discriminate analysis (PLS-DA) scores plot of the metabolite concentrations of *P. fluorescens* CHA0 (black), CHA19 (red), SCV (green) and WS (blue). Each data point represents a single extract and the position determined as a linear combination of 32 metabolite concentrations obtained from the ¹H NMR spectra. The four strains could be separated along three components. (below) Killing curves of *P. fluorescens* CHA0 (open squares), CHA19 (filled squares), SCV (closed circles) and WS (open circles). Biofilms of each of the strains were exposed to a series of metal concentrations for 4 h followed by viable cell counting. Shown are the log-killing values, which are the number of cells killed following the exposure to metal. Error bars represent standard deviation calculated from four replicates. Average cell counts for the initial unexposed controls were 5.38 ± 0.47 , 4.89 ± 0.45 , 5.23 ± 0.39 , 5.17 ± 0.47 log₁₀ cfu peg⁻¹ for CHA0, CHA19, SCV and WS respectively. Reprinted with permission from reference.²⁷¹

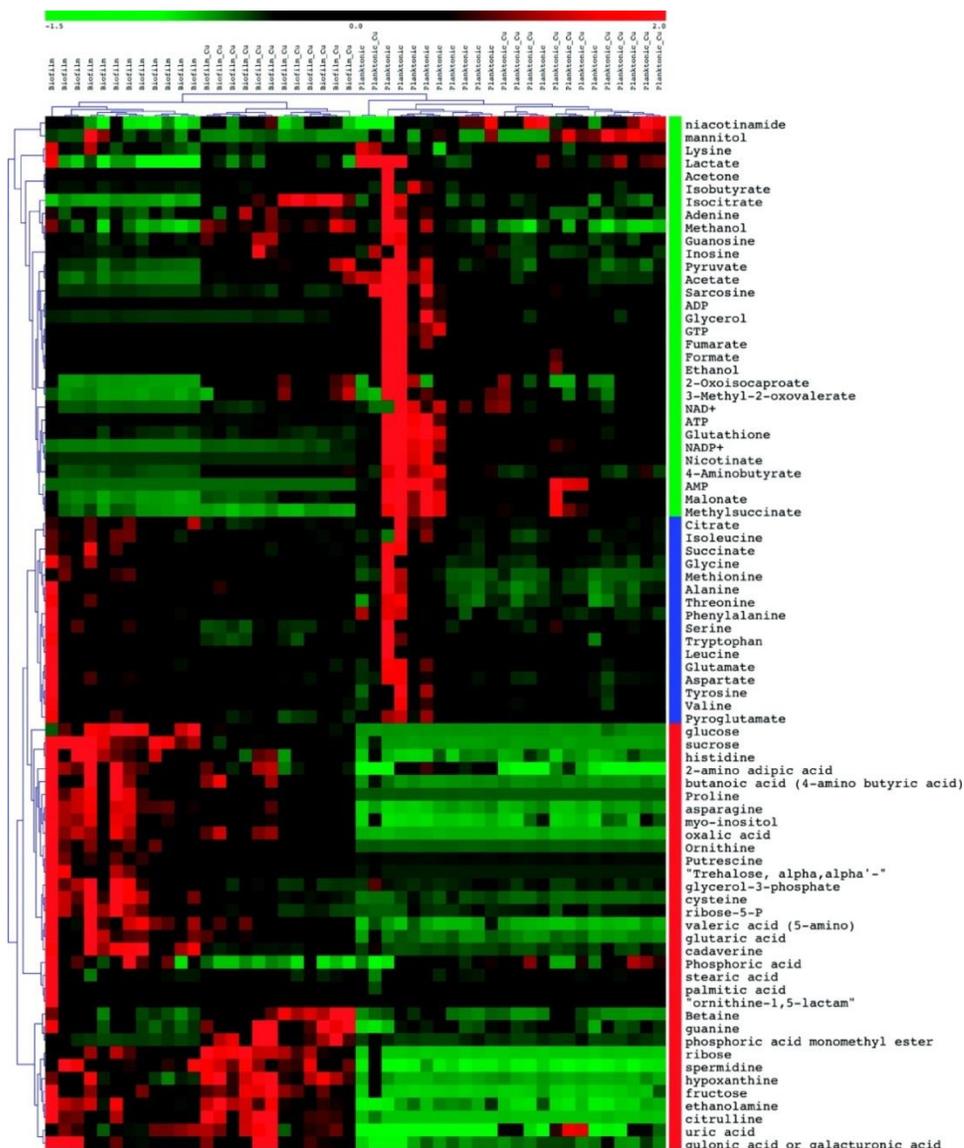


Figure 2.14 Hierarchical clustering analysis of control and copper exposed biofilm and planktonic culture metabolite concentrations. Identified metabolites were analyzed for any apparent patterns. Clustering was performed using Pearson correlation as the distance metric. The cluster tree shows how the samples and metabolites divide. Across the top are the samples, labeled by their class and along the side are the metabolites. These were colored according to the grouping pattern that they showed: green metabolites were only changing in the planktonic cultures, red only in biofilms and blue were changing in both cultures in response to copper exposure. Reprinted with permission from reference.¹⁶⁵

Demonstrating a difference between planktonic cells and biofilms by NMR is an important step towards an application in drug discovery. Does a chemical lead inhibit metabolic pathways associated with biofilm formation? Does drug treatment result in a metabolome more similar to planktonic cells despite conditions that induce biofilm formation? Comparing the metabolome of planktonic cells with and without drug treatment may provide an answer to these key questions. Furthermore, the comparative analysis of planktonic cells and biofilms has already identified changes in the activity of specific cellular process correlated with biofilm formation. Thus, proteins essential to exopolysaccharide production and the response to oxidative stress may be novel targets for disrupting biofilm formation. Additionally, the metabolome differences between planktonic cells and biofilms can be used as a diagnostic tool and to design treatments. As an illustration, Hall-Stoodley *et al.* (2006) describe the application of microbiological culture, polymerase chain reaction (PCR)-based diagnostics, direct microscopic examination, fluorescence *in situ* hybridization, and immunostaining to characterize middle-ear mucosa biopsy specimens for biofilm morphology.²⁷² The authors concluded that chronic otitis media in children that require tympanostomy tube placement is biofilm related. NMR metabolomics could provide a similar analysis of biopsy samples. The treatment of biofilm infections that have colonized on medical implants often requires the removal of the infected device.¹⁶⁶ Investigating changes in bacterial metabolomes in response to the different surfaces (metals, polymers, *etc.*) may contribute to the development of novel materials resistant to biofilm formation.^{273,274} Coating or embedding medical devices with antibiotics is a common approach to prevent biofilm

infections, but the over-use of antibiotics incurs the risk of inducing the rapid development of resistance.^{275,276}

Besides characterizing cellular differences through metabolomics, NMR can also be used to explore the overall structure and function of biofilms. Vogt *et al.* (2000) used NMR to describe differences in metabolite diffusion within a biofilm.²⁷⁷ PFG-NMR was used to measure diffusion coefficients for slowly moving water and other components in a *P. aeruginosa* biofilm. Five groups of components including water, glycerol, and polysaccharides, were observed with diffusion coefficients ranging from 1.8×10^{-9} to 5×10^{-13} ($\text{m}^2 \text{s}^{-1}$) that indicate locations in the biofilm pores or the extracellular polymeric substance (EPS). Correspondingly, the complicated structure of a biofilm is a major obstacle to a successful treatment with antibiotics. As the above NMR experiment infers, a biofilm is a diffusion barrier that hinders the infusion and dispersion of antibiotics within a biofilm.²⁷⁸ This also infers that the *in vivo* activity of a chemical lead is determined by both its efficient dispersion throughout the biofilm in addition to its intrinsic inhibitor activity.

A recent study by Rogers *et al.* (2010) analyzed the synergistic activity of 2-aminoimidazole-derived compounds, a new class of anti-biofilm agents that disperse biofilms.²⁷⁹ Combining a 2-aminoimidazole-derived compound with known antibiotics resulted in a 2- to 8-fold reduction in MICs against biofilms of *P. aeruginosa*, *Acinetobacter baumannii*, *Bordetella bronchiseptica*, and *S. aureus*. Importantly, the anti-biofilm agent actually resensitized a multi-drug resistant strain of *S. aureus* (MRSA). Similarly, Walencka *et al.* (2007) observed a synergy between salvipisone and

aethiopinone from *Salvia sclarea* Hairy Roots with β -lactam antibiotics. An improved activity was observed against MRSA and multi-resistant *S. epidermidis* (MRSE).²⁸⁰ Salvipisone and aethiopinone are postulated to function by altering cell surface hydrophobicity and cell wall/membrane permeability.

NMR metabolomics has also been used to investigate the mechanism by which *Staphylococcus epidermidis* and *S. aureus* respond to a diverse set of environmental signals to induce the planktonic to biofilm transition. Bacteria have been shown to form biofilms in response to variations in ethanol, oleic acid, glucose, UDP-N-acetylglucosamine, sub-inhibitory concentrations of some antibiotics, anaerobic conditions, Fe limitation, high osmolarity, and high temperature. Instead of numerous distinct signaling pathways, a series of detailed NMR and molecular biology experiments have demonstrated the presence of a single flexible metabolic signaling pathway centered on the tricarboxylic acid (TCA) cycle.^{141,221,281} 1D ¹H NMR combined with OPLS-DA was used to compare the metabolome of wild-type *S. epidermidis* 1457 and an aconitase mutant strain 1457-*acnA::tetM* under various environmental stressors known to induce biofilm formation (Figure 2.10). The change in the metabolome of wild-type *S. epidermidis* 1457 in the presence of 4% ethanol, 2% glucose, Fe-limitation and 0.06 g/mL of tetracycline was shown to be essentially identical to the aconitase mutant with an inactive TCA cycle. No change in the metabolome was observed for the aconitase mutant with or without the addition of environmental stress factors. 2D ¹H, ¹³C HSQC experiments combined with bacteria grown in the presence of ¹³C-glucose were used to generate a detailed analysis of the changes in the *S. epidermidis* metabolome (Figure

2.11). Consistent with the global changes in the metabolome, wild-type *S. epidermidis* in the presence of the environmental stressors induced the down-regulation of metabolites associated with the TCA cycle and the up-regulation of metabolites related to PIA production. These results suggest that biofilm formation is regulated by the activity of the TCA cycle. Inactivating the TCA cycle allows the shuttling of key metabolites into PIA production, which is generally necessary for biofilm formation. Similarly, White *et al.* (2011) compared the metabolome of wild-type *Salmonella* and a CsgD deletion mutant that prevents production of an extracellular matrix.²⁸² GC-MS and NMR were used to analyze the metabolome along with an analysis of gene expression. Metabolites associated with glucogenesis and major osmoprotectants were up-regulated in wild-type *Salmonella*, whereas; metabolites associated with the TCA cycle were up-regulated in the mutant. Again, this is consistent with the TCA cycle activity regulating biofilm formation. These results suggest that agonists of the TCA cycle would interfere with biofilm formation. Diets or drugs that modulate the nutrient environment may be an approach to prevent bacterial biofilm. As an example, iron-limitation down-regulates the TCA cycle and induces biofilm formation.²⁸³ The major source of morbidity and mortality in cystic fibrosis patients is *P. aeruginosa* biofilms formed in the lung. High iron concentrations inhibit *P. aeruginosa* biofilms, where chelated sources of iron combined with antibiotics hold promise as a treatment for cystic fibrosis.^{284,285}

Metabolomics can also be used to construct metabolic pathways, with contributions from proteomics and genomics information. Liebeke *et al.* (2011) provided a time-resolved analysis of *S. aureus* during the transition from exponential growth to

glucose starvation. The activity of more than 500 proteins and the concentration of 94 metabolites were followed. ^1H NMR was used for the quantification of compounds in the media before inoculation, and at defined time points during cell growth. Intracellular metabolites were measured by gas/liquid chromatography–mass spectrometry. In general, changes in the metabolome correlated with changes in the proteome, where the metabolome displayed a larger dynamic range. The most dramatic changes were observed for amino acids. During initial cell growth, glycolysis and protein-synthesis were highly active, but as glucose was exhausted gluconeogenesis and the TCA cycle were activated. Again, this is consistent with the TCA cycle activity regulating biofilm formation due to variations in glucose concentrations.⁹⁴ Metabolomics is a valuable approach to characterize the state of a system. Nevertheless, incorporating additional complementary data, such as proteomics, significantly enhances the reliability of the information. Observing a correlation between a metabolite's concentration and a protein's expression level further substantiates the importance and relevance of the protein to the system, such as biofilm formation. From a drug discovery perspective, this provides substantial corroboration of a potentially new drug discovery target.

2.9 Conclusion and future perspective

NMR metabolomics can be used to characterize different cell phenotypes, to investigate the underlying biology of biofilm formation, to explore the impact of various environmental stress factors on cell biology, to analyze the effect of gene mutations, to investigate the spatial and temporal structure of a biofilm, and even generate a three-

dimensional image. Notably, the NMR metabolomics methodologies used to study the biochemistry of bacterial biofilms are directly applicable to a drug discovery effort. NMR metabolomics has been used to identify disease biomarkers and diagnose a disease, to screen for drugs, to evaluate drug activity and toxicity, to identify new therapeutic targets and design new treatments. While the application of NMR for the analysis of changes in the metabolomics is a relatively new endeavor, the technique has already made some significant contributions to our understanding of bacterial biofilms. More importantly, NMR metabolomics holds great promise to significantly contribute to the diagnosis and treatment of biofilm related diseases, where it may play an important role in personalized medicine.

NMR-based metabolomics is a relatively new technology for systems biology and, correspondingly, has only had a limited use in the investigation of bacterial biofilms. This review summarized the majority of these studies. Nevertheless, the application of NMR-based metabolomics to the investigation of bacterial biofilms is only limited by the creativity of the scientific community. For the near future, the further development of the technology will be a primary focus. While NMR-based metabolomics is straight-forward in concept, there are numerous practical considerations that can severely complicate the routine application of the technique. A primary issue is our incomplete knowledge of the metabolome; extensive effort is still required to characterize the metabolome from all organisms and populate databases with reference NMR and MS spectra. Similarly, while some progress has been made, a metabolomics software package is still needed that automates and standardizes the processing of metabolomics data, chemometric analysis

and model validation, and metabolite identification. Also, the efficient and accurate extraction of metabolites from cell lysates requires continued optimization, and varies depending on the organism under investigation. Developing these protocols is extremely critical to the future success of metabolomics. Additional methodology advancements that will benefit metabolomics include the routine integration of MS and NMR data, and the efficient quantitation of metabolite concentrations from MS and 2D NMR experiments.

Our understanding of bacteria cell biology is far from complete, where NMR-based metabolomics will be an invaluable addition to the study of biofilms and related processes such as programmed cell death, inter-/intra-species communication, and pathogenesis. The systematic analysis of the bacterial genome, transcriptome, proteome and metabolome will enable the construction of a detailed network to describe the regulatory and metabolic pathways associated with biofilm formation, progression, evolution and survivability. In addition to enhancing our basic understanding of bacterial biofilms, NMR-based metabolomics will be an invaluable tool for drug discovery, disease diagnosis and personalized medicine. NMR-based metabolomics can be used as part of a drug discovery screening protocol. Observing an induced change in the bacterial metabolome due to a drug treatment would further validate a chemical lead identified from standard high-throughput screens.^{286,287} In fact, Halouska *et al.* (2012) recently demonstrated how NMR-based metabolomics can be used to identify the *in vivo* mechanism of action of a chemical lead.¹⁴⁰ This addresses a major challenge with drug discovery, identifying new drugs with activity against novel therapeutic targets that also

avoid common mechanisms of resistance or toxicity. Similarly, characterizing the metabolome of various pathogenic bacterial strains with a correlation to drug susceptibility provides a means to personalize patient treatments. In essence, the identification of characteristic metabolites from the biofluids of infected patients would identify the bacterial strain and preferred form of treatment.

2.10 References

- (1) O'Toole, G.; Kaplan, H. B.; Kolter, R. *Annual Review of Microbiology* **2000**, *54*, 49.
- (2) Costerton, J. W. *J Ind Microbiol* **1995**, *15*, 137.
- (3) Costerton, J. W.; Lewandowski, Z.; Caldwell, D. E.; Korber, D. R.; Lappin-Scott, H. M. *Annu Rev Microbiol* **1995**, *49*, 711.
- (4) Simões, L. C.; Simões, M.; Vieira, M. J. *Appl Environ Microbiol.* **2007**, *73*, 9.
- (5) Nadell, C. D.; Xavier, J. B.; Foster, K. R. *FEMS Microbiol. Rev.* **2009**, *33*, 206.
- (6) Angst, E. C.; Navy, U., Ed. Washington, DC, 1923.
- (7) Henrici, A. T. *Journal of Bacteriology* **1933**, *1933*, 10.
- (8) Heukelekian, H.; Heller, A. *J Bacteriol.* **1940**, *40*, 12.
- (9) Stewart, P.; Franklin, M. *Nat Rev Micro* **2008**, *6*, 199.

- (10) Hall-Stoodley, L.; Costerton, J. W.; Stoodley, P. *Nat Rev Micro* **2004**, *2*, 95.
- (11) Davey, M. E.; O'Toole, G. A. *Microbiol Mol Biol Rev* **2000**, *64*, 847.
- (12) Hiroyuki, K. *The American Journal of Medicine* **1995**, *99*, 26s.
- (13) Klaus-Joerger, T.; Joerger, R.; Olsson, E.; Granqvist, C.-G. *Trends in Biotechnology* **2001**, *19*, 15.
- (14) van Loosdrecht, M. C. M.; Eikelboom, D.; Gjaltema, A.; Mulder, A.; Tjihuis, L.; Heijnen, J. J. *Water Science and Technology* **1995**, *32*, 35.
- (15) Lens, P.; Moran, A. P.; Mahony, T.; Stoodley, P.; O'Flaherty, V. *Mater. Corros.* **2005**, *56*, 55.
- (16) Hughes, J. M. *Emerg Infect Dis* **1998**, *4*, 360.
- (17) Gannon, J. C. Washington, DC, 2000.
- (18) Lewis, K. *Antimicrob Agents Chemother* **2001**, *45*, 999.
- (19) Davies, D. *Nat Rev Drug Discov* **2003**, *2*, 114.
- (20) Fischbach, M. A.; Walsh, C. T. *Science* **2009**, *325*, 1089.
- (21) Kennedy, A. D.; Otto, M.; Braughton, K. R.; Whitney, A. R.; Chen, L.; Mathema, B.; Mediavilla, J. R.; Byrne, K. A.; Parkins, L. D.; Tenover, F. C.; Kreiswirth, B. N.; Musser, J. M.; DeLeo, F. R. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 1327.
- (22) Kohn, W. G.; Collins, A. S.; Cleveland, J. L.; Harte, J. A.; Eklund, K. J.; Malvitz, D. M. *MMWR Recomm Rep* **2003**, *52*, 1.
- (23) Hetrick, E. M.; Schoenfisch, M. H. *Chem. Soc. Rev.* **2006**, *35*, 780.

- (24) Parsek, M. R.; Singh, P. K. *Annu. Rev. Microbiol.* **2003**, *57*, 677.
- (25) Savage, D. C. *Annu Rev Microbiol* **1977**, *31*, 107.
- (26) Cunningham, A. B.; Lennox, J. E.; Ross, R. J. 2001.
- (27) Marsh, P. D.; Moter, A.; Devine, D. A. *Periodontol 2000* **2011**, *55*, 16.
- (28) Nicolle, L. E. *Drugs Aging* **2005**, *22*, 627.
- (29) Campoccia, D.; Montanaro, L.; Baldassarri, L.; An, Y. H.; Arciola, C. R. *International Journal of Artificial Organs* **2005**, *28*, 1186.
- (30) Nishimura, S.; Tsurumoto, T.; Yonekura, A.; Adachi, K.; Shindo, H. *Journal of Orthopaedic Science* **2006**, *11*, 46.
- (31) LaPlante, K. L.; Woodmansee, S. *Antimicrob Agents Chemother* **2009**, *53*, 3880.
- (32) Padera, R. F. *Cardiovasc. Pathol.* **2006**, *15*, 264.
- (33) Vadyvaloo, V.; Otto, M. *International Journal of Artificial Organs* **2005**, *28*, 1069.
- (34) Sohail, M. R.; Uslan, D. Z.; Khan, A. H.; Friedman, P. A.; Hayes, D. L.; Wilson, W. R.; Steckelberg, J. M.; Stoner, S.; Baddour, L. M. *J Am Coll Cardiol* **2007**, *49*, 1851.
- (35) Cunha, B. A.; Eisenstein, L. E.; Hamid, N. S. *Heart & lung : the journal of critical care* **2006**, *35*, 207.
- (36) Peacock, S. J.; Curtis, N.; Berendt, A. R.; Bowler, I. C.; Winearls, C. G.; Maxwell, P. *J Hosp Infect* **1999**, *41*, 223.

- (37) Knudsen, A. M.; Rosdahl, V. T.; Espersen, F.; Frimodt-Moller, N.; Skinhoj, P.; Bentzon, M. W. *J Hosp Infect* **1993**, *23*, 123.
- (38) Ghanem, G. A.; Boktour, M.; Warneke, C.; Pham-Williams, T.; Kassis, C.; Bahna, P.; Aboufaycal, H.; Hachem, R.; Raad, I. *Medicine (Baltimore)* **2007**, *86*, 54.
- (39) Hall, R. E.; Ash, A. S.; Ghali, W. A.; Moskowitz, M. A. *Am J Cardiol* **1997**, *79*, 1680.
- (40) Simon, D.; Fischer, S.; Grossman, A.; Downer, C.; Hota, B.; Heroux, A.; Trenholme, G. *Clin Infect Dis* **2005**, *40*, 1108.
- (41) Holman, W. L.; Rayburn, B. K.; McGiffin, D. C.; Foley, B. A.; Benza, R. L.; Bourge, R. C.; Pinderski, L. J.; Kirklin, J. K. *Ann Thorac Surg* **2003**, *75*, S48.
- (42) Herrmann, M.; Weyand, M.; Greshake, B.; von Eiff, C.; Proctor, R. A.; Scheld, H. H.; Peters, G. *Circulation* **1997**, *95*, 814.
- (43) Fischer, S. A.; Trenholme, G. M.; Costanzo, M. R.; Piccione, W. *Infectious complications in left ventricular assist device recipients*, Section of Infectious Disease, Rush Medical College, Chicago, Illinois, USA, 1997.
- (44) Holman, W. L.; Park, S. J.; Long, J. W.; Weinberg, A.; Gupta, L.; Tierney, A. R.; Adamson, R. M.; Watson, J. D.; Raines, E. P.; Couper, G. S.; Pagani, F. D.; Burton, N. A.; Miller, L. W.; Naka, Y. *J Heart Lung Transplant* **2004**, *23*, 1359.
- (45) Mohamed, J. A.; Huang, D. B. *J Med Microbiol* **2007**, *56*, 1581.
- (46) Olmstead, S.; Meiss, D.; Ralston, J. In *Newsletter Spring 2003*; KLAIRE LABS 2003.
- (47) Amann, R. I.; Krumholz, L.; Stahl, D. A. *J Bacteriol* **1990**, *172*, 762.
- (48) Brayton, P. R.; Tamplin, M. L.; Huq, A.; Colwell, R. R. *Appl Environ Microbiol* **1987**, *53*, 2862.

- (49) Caldwell, D. E.; Kocber, D. R.; Lawrence, J. R. *J. Microbiol. Methods* **1992**, *1992*, 12.
- (50) Lawrence, J. R.; Korber, D. R.; Hoyle, B. D.; Costerton, J. W.; Caldwell, D. E. *J. Bacteriol.* **1991**, *1991*, 9.
- (51) O'Toole, G. A.; Gibbs, K. A.; Hager, P. W.; Phibbs, P. V., Jr.; Kolter, R. *J Bacteriol* **2000**, *182*, 425.
- (52) Stanley, P. M. *Can J Microbiol* **1983**, *29*, 1493.
- (53) Watnick, P. I.; Kolter, R. *Mol Microbiol* **1999**, *34*, 586.
- (54) Maira-Litran, T.; Kropec, A.; Abeygunawardana, C.; Joyce, J.; Mark, G., III; Goldmann, D. A.; Pier, G. B. *Infection and Immunity* **2002**, *70*, 4433.
- (55) Mack, D.; Fischer, W.; Krokotsch, A.; Leopold, K.; Hartmann, R.; Egge, H.; Laufs, R. *J Bacteriol FIELD Full Journal Title:Journal of bacteriology* **1996**, *178*, 175.
- (56) van Loosdrecht, M. C.; Lyklema, J.; Norde, W.; Zehnder, A. J. *Microbiol Rev* **1990**, *54*, 75.
- (57) Wickham, J. R.; Rice, C. V. *Solid State Nucl. Magn. Reson.* **2008**, *34*, 154.
- (58) DeBeer, D.; Stoodley, P.; Roe, F. L.; Lewandowski, Z. *Biotech. Bioeng.* **1994**, *1994*, 8.
- (59) Lawrence, J. R.; Neu, T. R. *Methods Enzymol* **1999**, *310*, 131.
- (60) Webb, J. S.; Givskov, M.; Kjelleberg, S. *Curr. Opin. Microbiol.* **2003**, *6*, 578.
- (61) Krasovec, R.; Jerman, I. *Med. Hypotheses* **2003**, *60*, 484.

- (62) Ilyina, T. S.; Romanova, Y. M.; Gintsburg, A. L. *Russ. J. Genet.* **2004**, *40*, 1189.
- (63) Stoodley, P.; Sauer, K.; Davies, D. G.; Costerton, J. W. *Annu. Rev. Microbiol.* **2002**, *56*, 187.
- (64) Tanaka, G.; Shigeta, M.; Komatsuzawa, H.; Sugai, M.; Suginaka, H.; Usui, T. *Chemotherapy (Basel)* **1999**, *45*, 28.
- (65) Ashby, M. J.; Neale, J. E.; Knott, S. J.; Critchley, I. A. *J. Antimicrob. Chemother.* **1994**, *33*, 443.
- (66) Wolfaardt, G. M.; Lawrence, J. R.; Robarts, R. D.; Caldwell, S. J.; Caldwell, D. E. *Appl Environ Microbiol* **1994**, *60*, 434.
- (67) Wolfaardt, G. M.; Lawrence, J. R.; Robarts, R. D.; Caldwell, D. E. *Can J Microbiol* **1994**, *40*, 331.
- (68) Boyd, A.; Chakrabarty, A. M. *Appl Environ Microbiol* **1994**, *60*, 2355.
- (69) Picioreanu, C.; van Loosdrecht, M. C. M.; Heijnen, J. J. *Water Science and Technology* **1999**, *39*, 8.
- (70) Hao, X.; Heijnen, J. J.; van Loosdrecht, M. C. M. *Biotechnology and Bioengineering* **2002**, *77*, 266.
- (71) Park, S.; Bae, W.; Rittmann, B. E. *Biotechnol Bioeng* **2010**, *105*, 1115.
- (72) Chaudhry, M.; Beg, S. *Chemical Engineering & Technology* **1998**, *21*, 701.
- (73) Nelson, Y. M.; Lion, L. W.; Shuler, M. L.; Ghiorse, W. C. *Environmental Science & Technology* **1996**, *30*, 2027.
- (74) van Loosdrecht, M. C. M.; Heijnen, J. J.; Eberl, H.; Kreft, J.; Picioreanu, C. *Antonie van Leeuwenhoek* **2002**, *81*, 245.

- (75) Klapper, I.; Dockery, J. *SIAM J. Appl. Math.* **2001**, *62*, 17.
- (76) Eberl, H.; Morgenroth, E.; Noguera, D.; Picioreanu, C.; Rittmann, B.; van Loosdrecht, M.; Wanner, O. *Mathematical modeling of biofilms*; IWA Publishing, 2006; Vol. 18.
- (77) Wang, Q.; Zhang, T. *Solid State Communications* **2010**, *150*, 1009.
- (78) Fujikawa, H.; Matsushita, M. *Journal of the Physical Society of Japan* **1989**, *58*, 4.
- (79) Hermanowicz, S. W. *Math Biosci* **2001**, *169*, 1.
- (80) Costerton, J. W.; Stewart, P. S.; Greenberg, E. P. *Science* **1999**, *284*, 1318.
- (81) O'Toole, G. A.; Pratt, L. A.; Watnick, P. I.; Newman, D. K.; Weaver, V. B.; Kolter, R. *Methods Enzymol.* **1999**, *310*, 91.
- (82) Pratt, L. A.; Kolter, R. *Curr. Opin. Microbiol.* **1999**, *2*, 598.
- (83) Boles, B. R.; Thoendel, M.; Roth, A. J.; Horswill, A. R. *PLoS One* **2010**, *5*, No pp. given.
- (84) Chagneau, C.; Saier, M. H., Jr. *J. Mol. Microbiol. Biotechnol.* **2004**, *8*, 177.
- (85) Kristich, C. J.; Nguyen, V. T.; Le, T.; Barnes, A. M. T.; Grindle, S.; Dunny, G. M. *Appl. Environ. Microbiol.* **2008**, *74*, 3377.
- (86) Li, J.; Wang, N. *PLoS One* **2011**, *6*, e21804.
- (87) Quoc, P. H. T.; Genevaux, P.; Pajunen, M.; Savilahti, H.; Georgopoulos, C.; Schrenzel, J.; Kelley, W. L. *Infect. Immun.* **2007**, *75*, 1079.
- (88) Richard, M. L.; Nobile, C. J.; Bruno, V. M.; Mitchell, A. P. *Eukaryotic Cell* **2005**, *4*, 1493.

- (89) Yoshida, A.; Kuramitsu, H. K. *Appl. Environ. Microbiol.* **2002**, *68*, 6283.
- (90) Beloin, C.; Ghigo, J.-M. *Trends Microbiol.* **2005**, *13*, 16.
- (91) O'Toole, G. A. *J. Bacteriol.* **2003**, *185*, 2687.
- (92) Knobloch, J. K.; Bartscht, K.; Sabottke, A.; Rohde, H.; Feucht, H. H.; Mack, D. *J Bacteriol* **2001**, *183*, 2624.
- (93) Campbell, I. M.; Crozier, D. N.; Pawagi, A. B.; Buivids, I. A. *J Clin Microbiol* **1983**, *18*, 408.
- (94) Mack, D.; Siemssen, N.; Laufs, R. *Infect Immun* **1992**, *60*, 2048.
- (95) Gerke, C.; Kraft, A.; Sussmuth, R.; Schweitzer, O.; Gotz, F. *J Biol Chem* **1998**, *273*, 18586.
- (96) Rachid, S.; Ohlsen, K.; Witte, W.; Hacker, J.; Ziebuhr, W. *Antimicrobial Agents and Chemotherapy* **2000**, *44*, 3357.
- (97) Cramton, S. E.; Ulrich, M.; Gotz, F.; Doring, G. *Infection and Immunity* **2001**, *69*, 4079.
- (98) Deighton, M.; Borland, R. *Infect Immun* **1993**, *61*, 4473.
- (99) Elci, S.; Atmaca, S.; Gul, K. *Cytobios* **1995**, *84*, 141.
- (100) Evans, E.; Brown, M. R.; Gilbert, P. *Microbiology* **1994**, *140 (Pt 1)*, 153.
- (101) Rachid, S.; Ohlsen, K.; Wallner, U.; Hacker, J.; Hecker, M.; Ziebuhr, W. *J Bacteriol* **2000**, *182*, 6824.
- (102) Rowley, G.; Spector, M.; Kormanec, J.; Roberts, M. *Nat. Rev. Microbiol.* **2006**, *4*, 383.

- (103) Chaturongakul, S.; Raengpradub, S.; Wiedmann, M.; Boor, K. J. *Trends Microbiol.* **2008**, *16*, 388.
- (104) Sachdeva, P.; Misra, R.; Tyagi, A. K.; Singh, Y. *Febs j.* **2010**, *277*, 605.
- (105) Senn, M. M.; Bischoff, M.; von Eiff, C.; Berger-Bachi, B. *J Bacteriol* **2005**, *187*, 7397.
- (106) Gill, S. R.; Fouts, D. E.; Archer, G. L.; Mongodin, E. F.; Deboy, R. T.; Ravel, J.; Paulsen, I. T.; Kolonay, J. F.; Brinkac, L.; Beanan, M.; Dodson, R. J.; Daugherty, S. C.; Madupu, R.; Angiuoli, S. V.; Durkin, A. S.; Haft, D. H.; Vamathevan, J.; Khouri, H.; Utterback, T.; Lee, C.; Dimitrov, G.; Jiang, L.; Qin, H.; Weidman, J.; Tran, K.; Kang, K.; Hance, I. R.; Nelson, K. E.; Fraser, C. M. *J Bacteriol* **2005**, *187*, 2426.
- (107) Stock, A. M.; Robinson, V. L.; Goudreau, P. N. *Annu Rev Biochem* **2000**, *69*, 183.
- (108) Li, M.; Villaruz, A. E.; Vadyvaloo, V.; Sturdevant, D. E.; Otto, M. *BMC Microbiol* **2008**, *8*, 4.
- (109) Miller, M. B.; Bassler, B. L. *Annu Rev Microbiol* **2001**, *55*, 165.
- (110) Jayaraman, A.; Wood, T. K. *Annu. Rev. Biomed. Eng.* **2008**, *10*, 145.
- (111) Sturme, M. H.; Kleerebezem, M.; Nakayama, J.; Akkermans, A. D.; Vaughn, E. E.; de Vos, W. M. *Antonie Van Leeuwenhoek* **2002**, *81*, 233.
- (112) Xavier, K. B.; Bassler, B. L. *Curr Opin Microbiol* **2003**, *6*, 191.
- (113) Bordi, C.; de Bentzmann, S. *Annals of Intensive Care* **2011**, *1*, 1.
- (114) Chauhan, N.; Calderone, R. *Infect Immun* **2008**, *76*, 4795.
- (115) Raffa, R. B.; Iannuzzo, J. R.; Levine, D. R.; Saeid, K. K.; Schwartz, R. C.; Sucic, N. T.; Terleckyj, O. D.; Young, J. M. *J Pharmacol Exp Ther* **2005**, *312*, 417.

- (116) Kleerebezem, M.; Quadri, L. E.; Kuipers, O. P.; de Vos, W. M. *Mol Microbiol* **1997**, *24*, 895.
- (117) Gotoh, Y.; Eguchi, Y.; Watanabe, T.; Okamoto, S.; Doi, A.; Utsumi, R. *Curr. Opin. Microbiol.* **2010**, *13*, 232.
- (118) Fuqua, C.; Parsek, M. R.; Greenberg, E. P. *Annu. Rev. Genet.* **2001**, *35*, 439.
- (119) Taga, M. E.; Bassler, B. L. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100*, 14549.
- (120) Passador, L.; Tucker, K. D.; Guertin, K. R.; Journet, M. P.; Kende, A. S.; Iglewski, B. H. *J Bacteriol* **1996**, *178*, 5995.
- (121) Schaefer, A. L.; Hanzelka, B. L.; Eberhard, A.; Greenberg, E. P. *J Bacteriol* **1996**, *178*, 2897.
- (122) Rasko, D. A.; Moreira, C. G.; Li, D. R.; Reading, N. C.; Ritchie, J. M.; Waldor, M. K.; Williams, N.; Taussig, R.; Wei, S.; Roth, M.; Hughes, D. T.; Huntley, J. F.; Fina, M. W.; Falck, J. R.; Sperandio, V. *Science (Washington, DC, U. S.)* **2008**, *321*, 1078.
- (123) Balaban, N.; Cirioni, O.; Giacometti, A.; Ghiselli, R.; Braunstein, J. B.; Silvestri, C.; Mocchegiani, F.; Saba, V.; Scalise, G. *Antimicrob. Agents Chemother.* **2007**, *51*, 2226.
- (124) Roy, V.; Smith, J. A. I.; Wang, J.; Stewart, J. E.; Bentley, W. E.; Sintim, H. *O. J. Am. Chem. Soc.* **2010**, *132*, 11141.
- (125) Roy, V.; Adams, B. L.; Bentley, W. E. *Enzyme Microb. Technol.* **2011**, *49*, 113.
- (126) Rui, F.; Marques, J. C.; Miller, S. T.; Maycock, C. D.; Xavier, K. B.; Ventura, M. R. *Bioorg. Med. Chem.* **2012**, *20*, 249.

- (127) Tsuchikama, K.; Lowery, C. A.; Janda, K. D. *J. Org. Chem.* **2011**, *76*, 6981.
- (128) Lowery, C. A.; McKenzie, K. M.; Qi, L.; Meijler, M. M.; Janda, K. D. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 2395.
- (129) Roy, V.; Smith, J. A.; Wang, J.; Stewart, J. E.; Bentley, W. E.; Sintim, H. O. *J Am Chem Soc* **2010**, *132*, 11141.
- (130) Rui, F.; Marques, J. C.; Miller, S. T.; Maycock, C. D.; Xavier, K. B.; Ventura, M. R. *Bioorg Med Chem* **2012**, *20*, 249.
- (131) Reo, N. V. *Drug Chem Toxicol* **2002**, *25*, 375.
- (132) Fiehn, O. *Plant Mol Biol* **2002**, *48*, 155.
- (133) Oliver, S. G.; Winson, M. K.; Kell, D. B.; Baganz, F. *Trends in Biotechnology* **1998**, *16*, 373.
- (134) Beckonert, O.; Keun, H. C.; Ebbels, T. M.; Bundy, J.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. *Nat Protoc* **2007**, *2*, 2692.
- (135) Rubakhin, S. S.; Romanova, E. V.; Nemes, P.; Sweedler, J. V. *Nat Methods* **2011**, *8*, S20.
- (136) Roos, V.; Klemm, P. *Infect Immun* **2006**, *74*, 3565.
- (137) Vilain, S.; Cosette, P.; Zimmerlin, I.; Dupont, J. P.; Junter, G. A.; Jouenne, T. *J Proteome Res* **2004**, *3*, 132.
- (138) Peregrin-Alvarez, J. M.; Sanford, C.; Parkinson, J. *GenomeBiology* **2009**, *10*, No pp. given.
- (139) Monds, R. D.; O'Toole, G. A.; American Society for Microbiology: 2008, p 105.

- (140) Halouska, S.; Fenton, R. J.; Barletta, R. G.; Powers, R. *ACS Chemical Biology* **2012**, *7*, 166.
- (141) Zhang, B.; Halouska, S.; Schiaffo, C. E.; Sadykov, M. R.; Somerville, G. A.; Powers, R. *J Proteome Res* **2011**, *10*, 3743.
- (142) Shanaiah, N.; Zhang, S.; Desilva, M. A.; Raftery, D. 2008, p 341.
- (143) Bino, R. J.; Hall, R. D.; Fiehn, O.; Kopka, J.; Saito, K.; Draper, J.; Nikolau, B. J.; Mendes, P.; Roessner-Tunali, U.; Beale, M. H.; Trethewey, R. N.; Lange, B. M.; Wurtele, E. S.; Sumner, L. W. *Trends Plant Sci* **2004**, *9*, 418.
- (144) Powers, R. *Magnetic Resonance in Chemistry* **2009**, *47*, S2.
- (145) Coen, M.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. *Chem Res Toxicol* **2008**, *21*, 9.
- (146) Vinaixa, M.; Angel, R. M.; Rull, A.; Beltran, R.; Blade, C.; Brezmes, J.; Canellas, N.; Joven, J.; Correig, X. *J. Proteome Res.* **2010**, *9*, 2527.
- (147) Griffin, J. L.; Atherton, H.; Shockcor, J.; Atzori, L. *Nat. Rev. Cardiol.* **2011**, *8*, 630.
- (148) Li, H.; Wang, L.; Yan, X.; Liu, Q.; Yu, C.; Wei, H.; Li, Y.; Zhang, X.; He, F.; Jiang, Y. *J. Proteome Res.* **2011**, *10*, 2797.
- (149) McClay, J. L.; Adkins, D. E.; Isern, N. G.; O'Connell, T. M.; Wooten, J. B.; Zedler, B. K.; Dasika, M. S.; Webb, B. T.; Webb-Robertson, B.-J.; Pounds, J. G.; Murrelle, E. L.; Leppert, M. F.; van, d. O. E. J. C. G. *J. Proteome Res.* **2010**, *9*, 3083.
- (150) Sofia, M.; Maniscalco, M.; de, L. G.; Paris, D.; Melck, D.; Motta, A. *J. Biomed. Biotechnol.* **2011**, 403260.
- (151) MacIntyre, D. A.; Jimenez, B.; Lewintre, E. J.; Martin, C. R.; Schaefer, H.; Ballesteros, C. G.; Mayans, J. R.; Spraul, M.; Garcia-Conde, J.; Pineda-Lucena, A. *Leukemia* **2010**, *24*, 788.

- (152) Carrola, J.; Rocha, C. M.; Barros, A. S.; Gil, A. M.; Goodfellow, B. J.; Carreira, I. M.; Bernardo, J.; Gomes, A.; Sousa, V.; Carvalho, L.; Duarte, I. F. *J. Proteome Res.* **2011**, *10*, 221.
- (153) Napoli, C.; Sperandio, N.; Lawlor, R. T.; Scarpa, A.; Molinari, H.; Assfalg, M. *J. Proteome Res.*, Ahead of Print.
- (154) Bertini, I.; Cacciatore, S.; Jensen, B. V.; Schou, J. V.; Johansen, J. S.; Kruhoffer, M.; Luchinat, C.; Nielsen, D. L.; Turano, P. *Cancer Res.* **2012**, *72*, 356.
- (155) Kork, F.; Holthues, J.; Hellweg, R.; Jankowski, V.; Tepel, M.; Oehring, R.; Heuser, I.; Bierbrauer, J.; Peters, O.; Schlattmann, P.; Zidek, W.; Jankowski, J. *Curr. Alzheimer Res.* **2009**, *6*, 519.
- (156) Sinclair, A. J.; Viant, M. R.; Ball, A. K.; Burdon, M. A.; Walker, E. A.; Stewart, P. M.; Rauz, S.; Young, S. P. *NMR Biomed.* **2010**, *23*, 123.
- (157) Blasco, H.; Corcia, P.; Moreau, C.; Veau, S.; Fournier, C.; Vourc'h, P.; Emond, P.; Gordon, P.; Pradat, P.-F.; Praline, J.; Devos, D.; Nadal-Desbarats, L.; Andres, C. R. *PLoS One* **2010**, *5*, No pp. given.
- (158) Stoop, M. P.; Coulier, L.; Rosenling, T.; Shi, S.; Smolinska, A. M.; Buydens, L.; Ampt, K.; Stingl, C.; Dane, A.; Muilwijk, B.; Luitwieler, R. L.; Silievis, S. P. A. E.; Hintzen, R. Q.; Bischoff, R.; Wijmenga, S. S.; Hankemeier, T.; van, G. A. J.; Luiders, T. M. *Mol. Cell. Proteomics* **2010**, *9*, 2063.
- (159) Tiziani, S.; Kang, Y.; Choi, J. S.; Roberts, W.; Paternostro, G. *Nat. Commun.* **2011**, *2*, 1562/1.
- (160) Lindon, J. C.; Keun, H. C.; Ebbels, T. M. D.; Pearce, J. M. T.; Holmes, E.; Nicholson, J. K. *Pharmacogenomics* **2005**, *6*, 691.
- (161) Robertson, D. G.; Reily, M. D.; Baker, J. D. *J. Proteome Res.* **2007**, *6*, 526.
- (162) Wishart, D. S. *Drugs R&D* **2008**, *9*, 307.

(163) Fernie, A. R.; Trethewey, R. N.; Krotzky, A. J.; Willmitzer, L. *Nat Rev Mol Cell Biol* **2004**, *5*, 763.

(164) Ellis, D. I.; Dunn, W. B.; Griffin, J. L.; Allwood, J. W.; Goodacre, R. *Pharmacogenomics* **2007**, *8*, 1243.

(165) Booth, S. C.; Workentine, M. L.; Wen, J.; Shaykhutdinov, R.; Vogel, H. J.; Ceri, H.; Turner, R. J.; Weljie, A. M. *J Proteome Res* **2011**, *10*, 3190.

(166) Gjersing, E. L.; Herberg, J. L.; Horn, J.; Schaldach, C. M.; Maxwell, R. S. *Anal Chem* **2007**, *79*, 8037.

(167) Zulak, K. G.; Weljie, A. M.; Vogel, H. J.; Facchini, P. J. *BMC Plant Biol* **2008**, *8*, 5.

(168) Pears, M. R.; Cooper, J. D.; Mitchison, H. M.; Mortishire-Smith, R. J.; Pearce, D. A.; Griffin, J. L. *J Biol Chem* **2005**, *280*, 42508.

(169) Sadykov, M. R.; Zhang, B.; Halouska, S.; Nelson, J. L.; Kreimer, L. W.; Zhu, Y.; Powers, R.; Somerville, G. A. *Journal of Biological Chemistry* **2010**, *285*, 36616.

(170) Kjeldahl, K.; Bro, R. *Journal of chemometrics* **2010**, *24*, 558.

(171) Kumar, A.; Bhardwaj, A. *Biomed. Mater. (Bristol, U. K.)* **2008**, *3*, 034008/1.

(172) Fogue, P.; Halouska, S.; Werth, M.; Xu, K.; Harris, S.; Powers, R. *Journal of Proteome Research* **2006**, *5*, 1916.

(173) Bolten, C. J.; Kiefer, P.; Letisse, F.; Portais, J. C.; Wittmann, C. *Anal Chem* **2007**, *79*, 3843.

(174) Faijes, M.; Mars, A. E.; Smid, E. J. *Microb Cell Fact* **2007**, *6*, 27.

(175) Maharjan, R. P.; Ferenci, T. *Anal Biochem* **2003**, *313*, 145.

- (176) Canelas, A.; Ras, C.; ten Pierick, A.; van Dam, J.; Heijnen, J.; van Gulik, W. *Metabolomics* **2008**, *4*, 226.
- (177) Wu, X. H.; Yu, H. L.; Ba, Z. Y.; Chen, J. Y.; Sun, H. G.; Han, B. Z. *Biotechnol J* **2010**, *5*, 75.
- (178) Tredwell, G. D.; Edwards-Jones, B.; Leak, D. J.; Bundy, J. G. *PLoS ONE* **2011**, *6*, e16286.
- (179) Rabinowitz, J. D. *Expert Rev Proteomics* **2007**, *4*, 187.
- (180) Shaw, L.; Golonka, E.; Potempa, J.; Foster, S. J. *Microbiology* **2004**, *150*, 217.
- (181) Wittmann, C.; Kromer, J. O.; Kiefer, P.; Binz, T.; Heinzle, E. *Anal Biochem* **2004**, *327*, 135.
- (182) Weckwerth, W.; Wenzel, K.; Fiehn, O. *Proteomics* **2004**, *4*, 78.
- (183) Wu, H.; Southam, A. D.; Hines, A.; Viant, M. R. *Anal Biochem* **2008**, *372*, 204.
- (184) Bennett, B. D.; Kimball, E. H.; Gao, M.; Osterhout, R.; Van Dien, S. J.; Rabinowitz, J. D. *Nat Chem Biol* **2009**, *5*, 593.
- (185) Kell, D. B. *Current Opinion in Microbiology* **2004**, *7*, 296.
- (186) Blow, N. *Nature* **2008**, *455*, 697.
- (187) Wilson, I. D.; Plumb, R.; Granger, J.; Major, H.; Williams, R.; Lenz, E. M. *J Chromatogr B Analyt Technol Biomed Life Sci* **2005**, *817*, 67.
- (188) Freeman, R. *Magnetic resonance in chemistry and medicine*; Oxford University Press: Lindon, 2003.

- (189) Teng, Q.; Huang, W.; Collette, T. W.; Ekman, D. R.; Tan, C. *Metabolomics* **2009**, *5*, 199.
- (190) Moskau, D.; Ritcher, C.; Kovaks, H.; Salzmann, M.; Baselgia, L.; Haerberli, M.; Marek, D.; Schett, O. *Spectra Analyse* **2003**, *32*, 39.
- (191) Keun, H. C.; Beckonert, O.; Griffin, J. L.; Richter, C.; Moskau, D.; Lindon, J. C.; Nicholson, J. K. *Anal Chem* **2002**, *74*, 4588.
- (192) Carrieri, D.; McNeely, K.; De Roo, A. C.; Bennete, N.; Pelczer, I.; Dismukes, G. C. *Magn. Reson. Chem.* **2009**, *47*, S138.
- (193) Pan, Z.; Raftery, D. *Analytical and Bioanalytical Chemistry* **2007**, 387, 525.
- (194) Lanza, I. R.; Zhang, S.; Ward, L. E.; Karakelides, H.; Raftery, D.; Nair, K. S. *PLoS One* **2010**, *5*, No pp. given.
- (195) Pan, Z.; Raftery, D. *Anal. Bioanal. Chem.* **2007**, 387, 525.
- (196) Gu, H.; Pan, Z.; Xi, B.; Asiago, V.; Musselman, B.; Raftery, D. *Anal. Chim. Acta* **2011**, *686*, 57.
- (197) Atherton, H. J.; Bailey, N. J.; Zhang, W.; Taylor, J.; Major, H.; Shockcor, J.; Clarke, K.; Griffin, J. L. *Physiol. Genomics* **2006**, *27*, 178.
- (198) Wang, H.; Manicke, N. E.; Yang, Q.; Zheng, L.; Shi, R.; Cooks, R. G.; Ouyang, Z. *Anal. Chem. (Washington, DC, U. S.)* **2011**, *83*, 1197.
- (199) Weljie, A. M.; Newton, J.; Mercier, P.; Carlson, E.; Slupsky, C. M. *Anal Chem* **2006**, *78*, 4430.
- (200) Moestue, S.; Sitter, B.; Bathen, T. F.; Tessem, M.-B.; Gribbestad, I. S. *Curr. Top. Med. Chem. (Sharjah, United Arab Emirates)* **2011**, *11*, 2.

- (201) Sitter, B.; Bathen, T. F.; Singstad, T. E.; Fjosne, H. E.; Lundgren, S.; Halgunset, J.; Gribbestad, I. S. *NMR Biomed.* **2010**, *23*, 424.
- (202) Li, M.; Song, Y.; Cho, N.; Chang, J. M.; Koo, H. R.; Yi, A.; Kim, H.; Park, S.; Moon, W. K. *PLoS ONE* **2011**, *6*, e25563.
- (203) Portais, J. C.; Delort, A. M. *FEMS Microbiol Rev* **2002**, *26*, 375.
- (204) Zamboni, N.; Sauer, U. *Curr Opin Microbiol* **2009**, *12*, 553.
- (205) DeSilva, M. A.; Shanaiah, N.; Nagana Gowda, G. A.; Rosa-Perez, K.; Hanson, B. A.; Raftery, D. *Magn Reson Chem* **2009**, *47 Suppl 1*, S74.
- (206) Meiboom, S.; Gill, D. *Rev. Sci. Instrum.* **1958**, *29*, 4.
- (207) Kaiser, K. A.; Barding, G. A., Jr.; Larive, C. K. *Magn. Reson. Chem.* **2009**, *47*, S147.
- (208) Martineau, E.; Tea, I.; Loaec, G.; Giraudeau, P.; Akoka, S. *Anal. Bioanal. Chem.* **2011**, *401*, 2133.
- (209) William, S. In *Annual Reports on NMR Spectroscopy*; Webb, G. A., Ed.; Academic Press: 1999; Vol. Volume 38, p 289.
- (210) Beckonert, O.; Coen, M.; Keun, H. C.; Wang, Y.; Ebbels, T. M.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. *Nat Protoc* **2010**, *5*, 1019.
- (211) Cloarec, O.; Dumas, M.-E.; Craig, A.; Barton, R. H.; Trygg, J.; Hudson, J.; Blancher, C.; Gauguier, D.; Lindon, J. C.; Holmes, E.; Nicholson, J. *Analytical Chemistry* **2005**, *77*, 1282.
- (212) Crockford, D. J.; Holmes, E.; Lindon, J. C.; Plumb, R. S.; Zirah, S.; Bruce, S. J.; Rainville, P.; Stumpf, C. L.; Nicholson, J. K. *Analytical Chemistry* **2005**, *78*, 363.
- (213) Xi, Y.; de Ropp, J.; Viant, M.; Woodruff, D.; Yu, P. *Metabolomics* **2006**, *2*, 221.

- (214) Ludwig, C.; Ward, D. G.; Martin, A.; Viant, M. R.; Ismail, T.; Johnson, P. J.; Wakelam, M. J.; Gunther, U. L. *Magn Reson Chem* **2009**, *47 Suppl 1*, S68.
- (215) Ludwig, C.; Viant, M. R. *Phytochem. Anal.* **2010**, *21*, 22.
- (216) Aue, W. P.; Karhan, J.; Ernst, R. R. *J. Chem. Phys.* **1976**, *1976*, 2.
- (217) Fonville, J. M.; Maher, A. D.; Coen, M.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. *Anal. Chem. (Washington, DC, U. S.)* **2010**, *82*, 1811.
- (218) Palmer, A. I.; Cavanagh, J.; Wright, P.; Rance, M. *J. Magn. Reson. J. Magn. Reson.* **1991**, *93*, 20.
- (219) Peyrauda, R.; Kiefera, P.; Christena, P.; Massoub, S.; Portaisb, J.; Vorholta, J. *PNAS* **2009**, *106*, 6.
- (220) Jans, A. W.; Willem, R. *Eur J Biochem* **1988**, *174*, 67.
- (221) Sadykov, M. R.; Zhang, B.; Halouska, S.; Nelson, J. L.; Kreimer, L. W.; Zhu, Y.; Powers, R.; Somerville, G. A. *J. Biol. Chem.* **2010**, *285*, 36616.
- (222) Shanaiah, N.; Desilva, M. A.; Nagana Gowda, G. A.; Raftery, M. A.; Hainline, B. E.; Raftery, D. *Proc Natl Acad Sci U S A* **2007**, *104*, 11540.
- (223) Hu, K.; Westler, W. M.; Markley, J. L. *Journal of the American Chemical Society* **2011**, *133*, 1662.
- (224) Holmes, E.; Foxall, P. J. D.; Nicholson, J. K.; Neild, G. H.; Brown, S. M.; Beddell, C. R.; Sweatman, B. C.; Rahr, E.; Lindon, J. C.; et al. *Analytical Biochemistry* **1994**, *220*, 284.
- (225) Ross, A.; Schlotterbeck, G.; Klaus, W.; Senn, H. *Journal of Biomolecular NMR* **2000**, *16*, 139.
- (226) Anderson, P. E.; Reo, N. V.; DelRaso, N. J.; Doom, T. E.; Raymer, M. L. *Metabolomics* **2008**, *4*, 261.

- (227) Davis, R. A.; Charlton, A. J.; Godward, J.; Jones, S. A.; Harrison, M.; Wilson, J. C. *Chemom. Intell. Lab. Syst.* **2007**, *85*, 144.
- (228) De Meyer, T.; Sinnaeve, D.; Van Gasse, B.; Tsiporkova, E.; Rietzschel, E. R.; De Buyzere, M. L.; Gillebert, T. C.; Bekaert, S.; Martins, J. C.; Van Criekeing, W. *Anal. Chem. (Washington, DC, U. S.)* **2008**, *80*, 3783.
- (229) Anderson, P. E.; Mahle, D. A.; Doom, T. E.; Reo, N. V.; Del, R. N. J.; Raymer, M. L. *Metabolomics* **2011**, *7*, 179.
- (230) Halouska, S.; Powers, R. *Journal of Magnetic Resonance* **2006**, *178*, 88.
- (231) Wu, W.; Daszykowski, M.; Walczak, B.; Sweatman, B. C.; Connor, S. C.; Haselden, J. N.; Crowther, D. J.; Gill, R. W.; Lutz, M. W. *J. Chem. Inf. Model.* **2006**, *46*, 863.
- (232) Forshed, J.; Schuppe-Koistinen, I.; Jacobsson, S. P. *Anal. Chim. Acta* **2003**, *487*, 189.
- (233) Veselkov, K. A.; Lindon, J. C.; Ebbels, T. M. D.; Crockford, D.; Volynkin, V. V.; Holmes, E.; Davies, D. B.; Nicholson, J. K. *Anal. Chem. (Washington, DC, U. S.)* **2009**, *81*, 56.
- (234) Csenki, L.; Alm, E.; Torgrip, R. J. O.; Aaberg, K. M.; Nord, L. I.; Schuppe-Koistinen, I.; Lindberg, J. *Anal. Bioanal. Chem.* **2007**, *389*, 875.
- (235) Fogh, R. H.; Vranken, W. F.; Boucher, W.; Stevens, T. J.; Laue, E. D. *J Biomol NMR* **2006**, *36*, 147.
- (236) Pons, J. L.; Malliavin, T. E.; Delsuc, M. A. *J Biomol NMR* **1996**, *8*, 445.
- (237) van Beek, J. D. *J Magn Reson* **2007**, *187*, 19.
- (238) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J Biomol NMR* **1995**, *6*, 277.

(239) Lewis, I. A.; Schommer, S. C.; Markley, J. L. *Magn Reson Chem* **2009**, *47 Suppl 1*, S123.

(240) Wilson, M.; Reynolds, G.; Kauppinen, R. A.; Arvanitis, T. N.; Peet, A. C. *Magn Reson Med* **2011**, *65*, 1.

(241) Craig, A.; Cloarec, O.; Holmes, E.; Nicholson, J. K.; Lindon, J. C. *Anal Chem* **2006**, *78*, 2262.

(242) Tim, D. M.; Sinnaeve, D.; Van Gasse, B.; Rietzschel, E.-R.; De Buyzere, M. L.; Langlois, M. R.; Bekaert, S.; Martins, J. C.; Van Criekeing, W. *Anal. Bioanal. Chem.* **2010**, *398*, 1781.

(243) Oliver, F. In *The Handbook of Metabonomics and Metabolomics* Lindon, J. C., Nicholson, J. K., Holmes, E. , Ed.; Elsevier Science: Amsterdam, 2007, p 20.

(244) Lindon, J. C.; Nicholson, J. K.; Holmes, E. *The Handbook of Metabonomics and Metabolomics* 2007.

(245) Basant, A.; Rege, M.; Sharma, S.; Sonawat, H. M. *Malaria Journal* **2010**, *9*, 110.

(246) Westerhuis, J. A.; Hoefsloot, H. C. J.; Smit, S.; Vis, D. J.; Smilde, A. K.; van, V. E. J. J.; van, D. J. P. M.; van, D. F. A. *Metabolomics* **2008**, *4*, 81.

(247) Rubingh, C. M.; Bijlsma, S.; Derks, E. P. P. A.; Bobeldijk, I.; Verheij, E. R.; Kochhar, S.; Smilde, A. K. *Metabolomics* **2006**, *2*, 53.

(248) Keun, H. C.; Ebbels, T. M.; Antti, H.; Bollard, M. E.; Beckonert, O.; Schlotterbeck, G.; Senn, H.; Niederhauser, U.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. *Chem Res Toxicol* **2002**, *15*, 1380.

(249) Wang, T.; Shao, K.; Chu, Q.; Ren, Y.; Mu, Y.; Qu, L.; He, J.; Jin, C.; Xia, B. *BMC Bioinformatics* **2009**, *10*, 83.

- (250) Zhao, Q.; Stoyanova, R.; Du, S.; Sajda, P.; Brown, T. R. *Bioinformatics* **2006**, *22*, 2562.
- (251) Xia, J.; Psychogios, N.; Young, N.; Wishart, D. S. *Nucleic Acids Res* **2009**, *37*, W652.
- (252) Izquierdo-Garcia, J.; Rodriguez, I.; Kyriazis, A.; Villa, P.; Barreiro, P.; Desco, M.; Ruiz-Cabello, J. *BMC Bioinformatics* **2009**, *10*, 363.
- (253) Werth, M. T.; Halouska, S.; Shortridge, M. D.; Zhang, B.; Powers, R. *Analytical Biochemistry* **2010**, *399*, 58.
- (254) Efron, B.; Halloran, E.; Holmes, S. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 13429.
- (255) Retief, J. D. *Methods in Molecular Biology* **1999**, *132*, 16.
- (256) Shulaev, V. *Briefings Bioinf.* **2006**, *7*, 128.
- (257) Oberhardt, M. A.; Palsson, B. O.; Papin, J. A. *Mol Syst Biol* **2009**, *5*, 320.
- (258) Robinette, S. L.; Ajredini, R.; Rasheed, H.; Zeinomar, A.; Schroeder, F. C.; Dossey, A. T.; Edison, A. S. *Anal Chem* **2011**.
- (259) Vu, T. N.; Valkenborg, D.; Smets, K.; Verwaest, K. A.; Dommissie, R.; Lemiere, F.; Verschoren, A.; Goethals, B.; Laukens, K. *BMC Bioinformatics* **2011**, *12*, 405.
- (260) Mo, H.; Raftery, D. *Anal Chem* **2008**, *80*, 9835.
- (261) Craig, A.; Cloarec, O.; Holmes, E.; Nicholson, J. K.; Lindon, J. C. *Anal. Chem.* **2006**, *78*, 2262.
- (262) Xia, J.; Bjorndahl, T. C.; Tang, P.; Wishart, D. S. *BMC Bioinformatics* **2008**, *9*, 507.

(263) Cui, Q.; Lewis, I. A.; Hegeman, A. D.; Anderson, M. E.; Li, J.; Schulte, C. F.; Westler, W. M.; Eghbalnia, H. R.; Sussman, M. R.; Markley, J. L. *Nat Biotechnol* **2008**, *26*, 162.

(264) Ulrich, E. L.; Akutsu, H.; Doreleijers, J. F.; Harano, Y.; Ioannidis, Y. E.; Lin, J.; Livny, M.; Mading, S.; Maziuk, D.; Miller, Z.; Nakatani, E.; Schulte, C. F.; Tolmie, D. E.; Kent Wenger, R.; Yao, H.; Markley, J. L. *Nucleic Acids Res* **2008**, *36*, D402.

(265) Wishart, D. S.; Knox, C.; Guo, A. C.; Eisner, R.; Young, N.; Gautam, B.; Hau, D. D.; Psychogios, N.; Dong, E.; Bouatra, S.; Mandal, R.; Sinelnikov, I.; Xia, J.; Jia, L.; Cruz, J. A.; Lim, E.; Sobsey, C. A.; Shrivastava, S.; Huang, P.; Liu, P.; Fang, L.; Peng, J.; Fradette, R.; Cheng, D.; Tzur, D.; Clements, M.; Lewis, A.; De, S. A.; Zuniga, A.; Dawe, M.; Xiong, Y.; Clive, D.; Greiner, R.; Nazyrova, A.; Shaykhtudinov, R.; Li, L.; Vogel, H. J.; Forsythe, I. *Nucleic Acids Res.* **2009**, *37*, D603.

(266) Kanehisa, M.; Araki, M.; Goto, S.; Hattori, M.; Hirakawa, M.; Itoh, M.; Katayama, T.; Kawashima, S.; Okuda, S.; Tokimatsu, T.; Yamanishi, Y. *Nucleic Acids Res* **2008**, *36*, D480.

(267) Caspi, R.; Foerster, H.; Fulcher, C. A.; Kaipa, P.; Krummenacker, M.; Latendresse, M.; Paley, S.; Rhee, S. Y.; Shearer, A. G.; Tissier, C.; Walk, T. C.; Zhang, P.; Karp, P. D. *Nucleic Acids Res* **2008**, *36*, D623.

(268) Cline, M. S.; Smoot, M.; Cerami, E.; Kuchinsky, A.; Landys, N.; Workman, C.; Christmas, R.; Avila-Campilo, I.; Creech, M.; Gross, B.; Hanspers, K.; Isserlin, R.; Kelley, R.; Killcoyne, S.; Lotia, S.; Maere, S.; Morris, J.; Ono, K.; Pavlovic, V.; Pico, A. R.; Vailaya, A.; Wang, P. L.; Adler, A.; Conklin, B. R.; Hood, L.; Kuiper, M.; Sander, C.; Schmulevich, I.; Schwikowski, B.; Warner, G. J.; Ideker, T.; Bader, G. D. *Nat Protoc* **2007**, *2*, 2366.

(269) Viswanathan, G. A.; Seto, J.; Patil, S.; Nudelman, G.; Sealfon, S. C. *PLoS Comput Biol* **2008**, *4*, e16.

(270) Funahashi, A.; Jouraku, A.; Matsuoka, Y.; Kitano, H. *In Silico Biol* **2007**, *7*, S81.

(271) Workentine, M. L.; Harrison, J. J.; Weljie, A. M.; Tran, V. A.; Stenroos, P. U.; Tremaroli, V.; Vogel, H. J.; Cerl, H.; Turner, R. J. *Environ. Microbiol.* **2010**, *12*, 1565.

(272) Hall-Stoodley, L.; Hu, F. Z.; Gieseke, A.; Nistico, L.; Nguyen, D.; Hayes, J.; Forbes, M.; Greenberg, D. P.; Dice, B.; Burrows, A.; Wackym, P. A.; Stoodley, P.; Post, J. C.; Ehrlich, G. D.; Kerschner, J. E. *JAMA* **2006**, *296*, 202.

(273) Kim, Y. H.; An, E. S.; Song, B. K.; Kim, D. S.; Chelikani, R. *Biotechnol Lett* **2003**, *25*, 1521.

(274) Carlson, R. P.; Taffs, R.; Davison, W. M.; Stewart, P. S. *J Biomater Sci Polym Ed* **2008**, *19*, 1035.

(275) Khoo, X.; Grinstaff, M. W. *MRS Bull.* **2011**, *36*, 357.

(276) Shunmugaperumal, T. *Recent Pat. Drug Delivery Formulation* **2010**, *4*, 153.

(277) Vogt, M.; Flemming, H. C.; Veeman, W. S. *J Biotechnol* **2000**, *77*, 137.

(278) Hoyle, B. D.; Alcantara, J.; Costerton, J. W. *Antimicrob. Agents Chemother.* **1992**, *36*, 2054.

(279) Rogers, S. A.; Huigens, R. W., 3rd; Cavanagh, J.; Melander, C. *Antimicrob Agents Chemother* **2010**, *54*, 2112.

(280) Walencka, E.; Rozalska, S.; Wysokinska, H.; Rozalski, M.; Kuzma, L.; Rozalska, B. *Planta Med* **2007**, *73*, 545.

(281) Sadykov, M. R.; Olson, M. E.; Halouska, S.; Zhu, Y.; Fey, P. D.; Powers, R.; Somerville, G. A. *Journal of Bacteriology* **2008**, *130*, 7621.

(282) White, A. P.; Weljie, A. M.; Apel, D.; Zhang, P.; Shaykhutdinov, R.; Vogel, H. J.; Surette, M. G. *PLoS ONE* **2010**, *5*, e11814.

- (283) Theodore, T. S.; Schade, A. L. *J Gen Microbiol* **1965**, *40*, 385.
- (284) Musk, D. J., Jr.; Hergenrother, P. J. *J. Appl. Microbiol.* **2008**, *105*, 380.
- (285) Moreau-Marquis, S.; O'Toole, G. A.; Stanton, B. A. *Am. J. Respir. Cell Mol. Biol.* **2009**, *41*, 305.
- (286) Persidis, A. *Nature Biotechnology* **1998**, *16*, 488.
- (287) Sittampalam, G. S.; Kahl, S. D.; Janzen, W. P. *Current Opinion in Chemical Biology* **1997**, *1*, 384.

CHAPTER 3

REVISITING PROTOCOLS FOR THE NMR ANALYSIS OF BACTERIAL CENTRAL 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 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 ¹³C- and ¹⁵N-isotope labeling to enhance specific regions of the metabolome.^{25,26} Importantly, the choice of ¹³C- or ¹⁵N-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 (≥ 1 to 5 μ M) 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 plants³³ 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) ^1H 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) ^1H , ^{13}C Heteronuclear Single Quantum Coherence (HSQC) or ^1H , ^1H 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.

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.

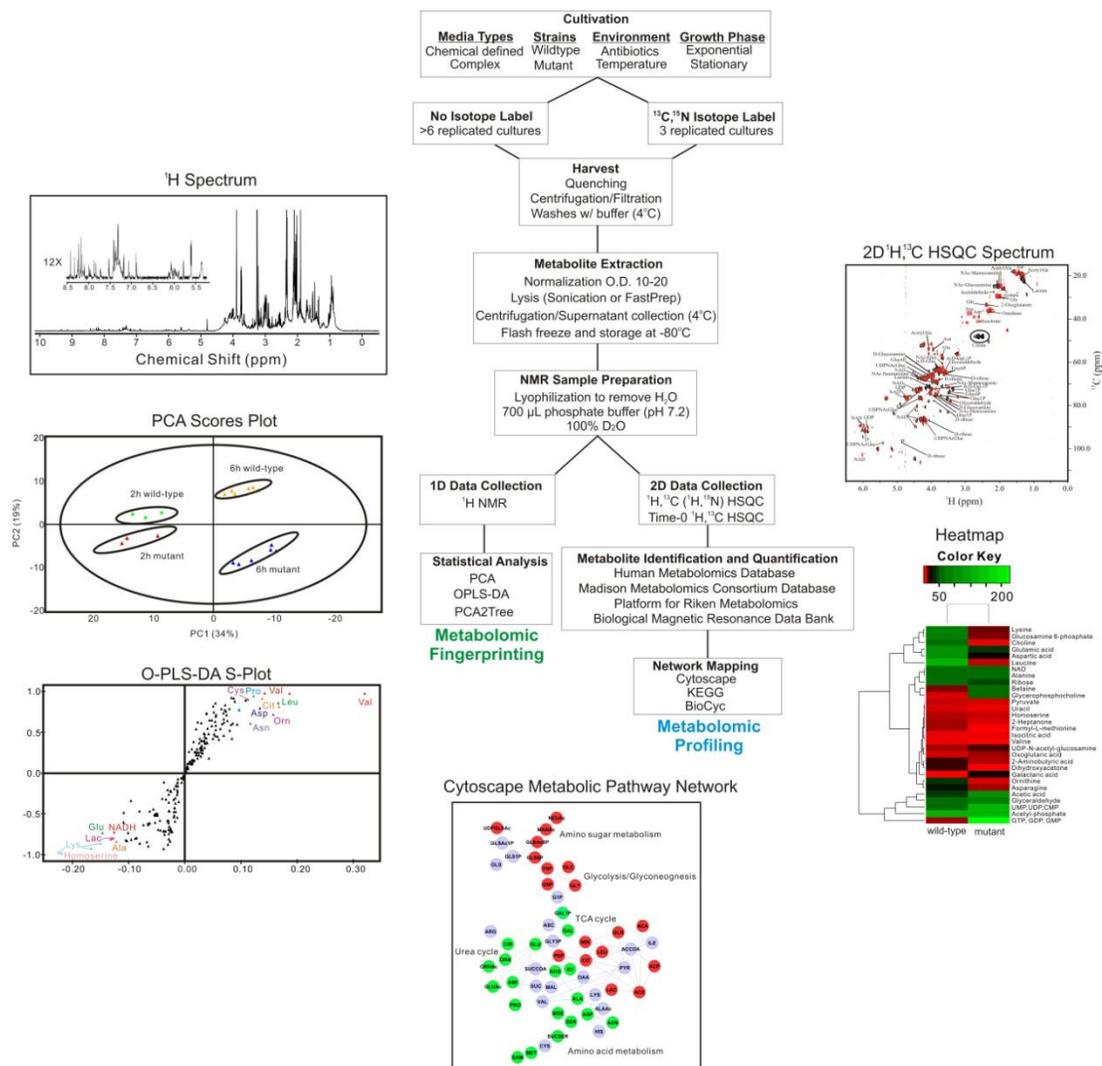


Figure 3.1 A flow chart of our protocol used for the NMR analysis of bacterial metabolomes.

3.2.1 Identify the 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 targets⁴⁹ 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 Minimize 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 ^1H NMR or 2D ^1H , ^{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 (≥ 1 to 5 M) 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 $^{13}\text{C}_6$ -glucose or ~10-15 mg/L of a targeted metabolite like ^{13}C -D-alanine in order to acquire a 2D ^1H , ^{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 (MIC₅₀), 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 MIC₅₀ 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 MIC₅₀ 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°C to ~ 0° 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°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°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 contaminants, 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 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.

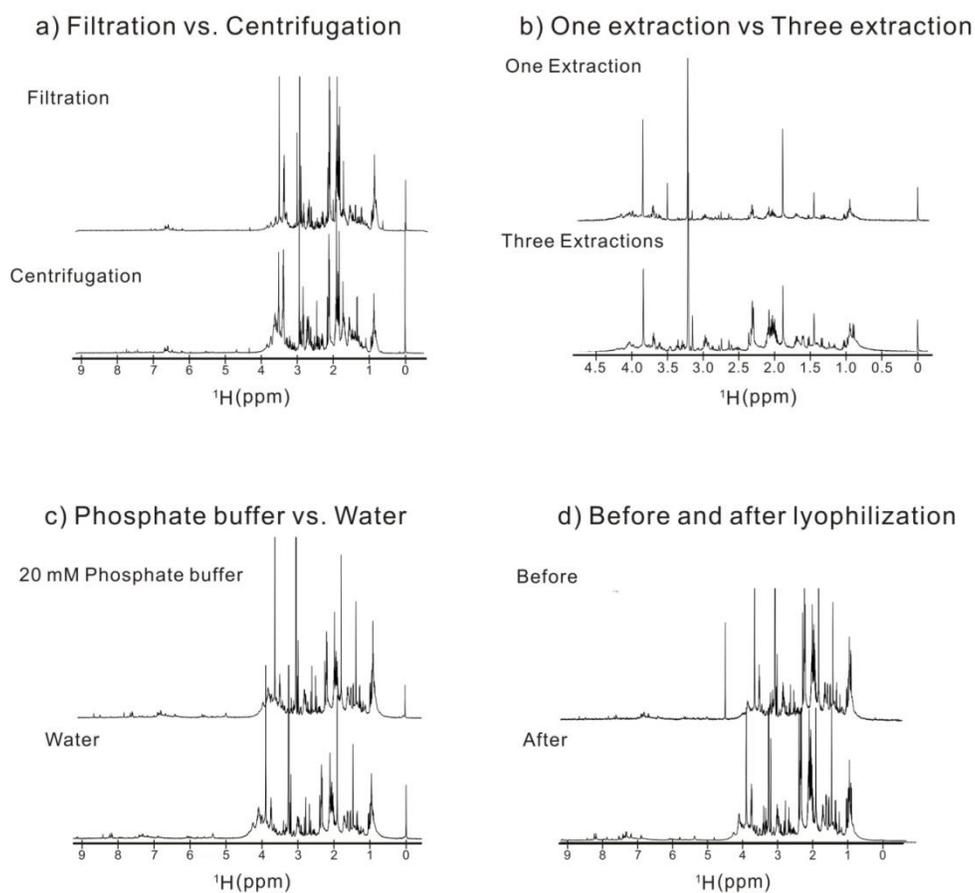


Figure 3.2 Illustrations of the impact of a) filtration and centrifugation, b) number of extraction steps, c) type of wash buffer, and d) lyophilization on the composition of the metabolome.

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 L of a phosphate buffer in D₂O 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°C freezer or directly lyophilized over-night.

3.3.4 Isotopic labeling of metabolites

Uniform or selective isotopic labeling with ¹³C/¹⁵N has often been used to boost the sensitivity of carbon or nitrogen due to their low natural abundance in protein

structural determination by NMR.^{66,67} The use of stable isotope tracing in metabolic pathway studies is dated back to 1935.⁶⁸ And there is no evidence that $^{13}\text{C}/^{15}\text{N}$ isotopic labelling has an observable effect on the physiology of bacteria.⁶⁹ Hence, stable isotope labeling combines the analytical development and biological requirement and demonstrates large potentials in the metabolomic study by handful applications.⁷⁰⁻⁷³ The uniform labeling by ^{13}C glucose is a common strategy in NMR-based metabolomics.⁷⁴ It often employs chemical defined medium (CDM) in which glucose as the sole carbon source is used to completely label the metabolome. CDM is not ideal for many physiological studies in bacteria because it has the drawbacks of less reproducibility and less association with the growth conditions in the host.⁷⁵ Therefore random labeling of ^{13}C isotope using ^{13}C glucose within a complex medium can also be performed with a good signal-to-noise ratio (Figure 3.3a). ^{13}C glucose should be added to the medium which is free of dextrose. However, there is a concern that if the ^{13}C glucose should be used in the overnight culture, or in another word, if the labeling of overnight culture makes a difference on the labeling efficiency. In Figure 3.3b and 3.3d, there was no significant difference between the two overnight cultures. In Figure 3.3a and 3.3c, as a negative control experiment, ^{12}C glucose was used for the post-exponential growth using the overnight cultures. And very little difference was found. In conclusion, the small amount of overnight culture has no significant effect on the labeling efficiency.

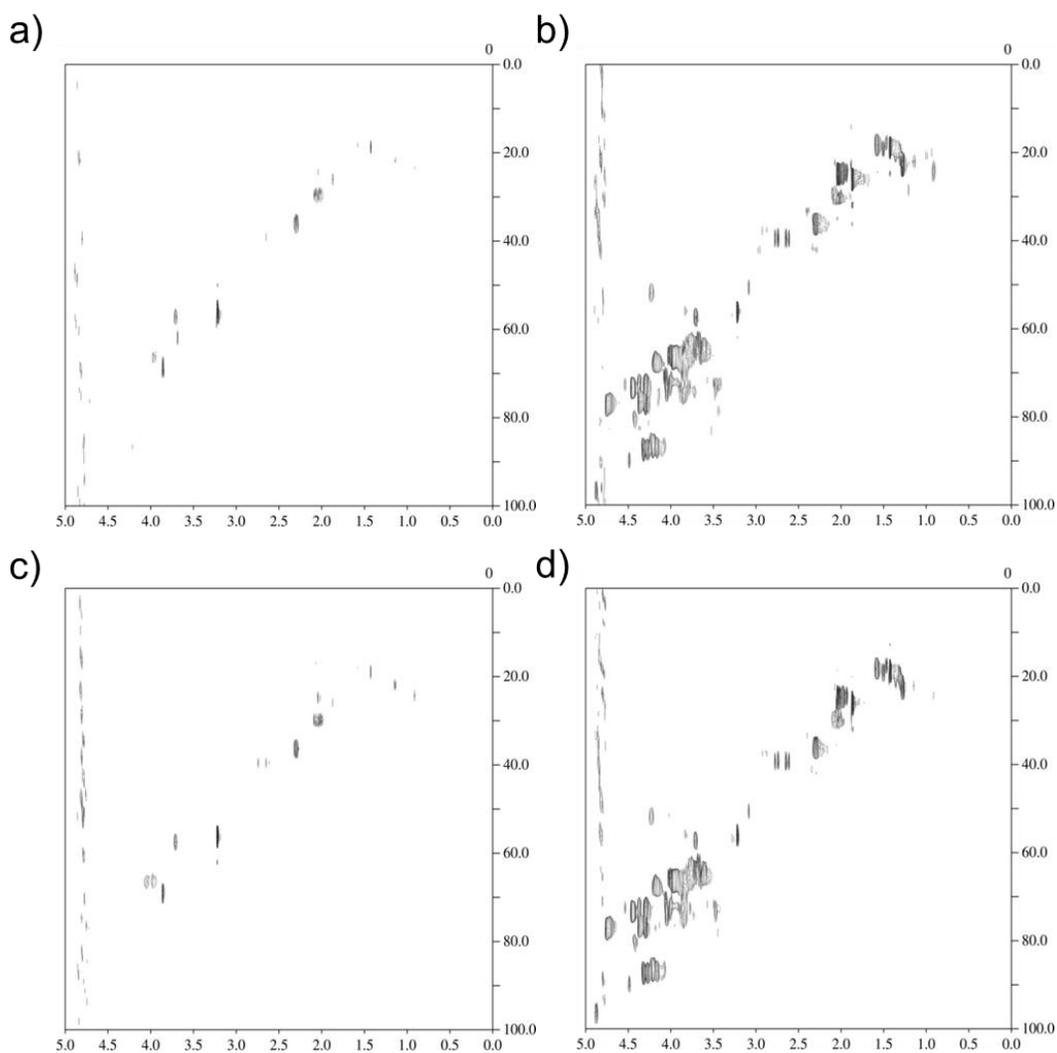


Figure 3.3 Illustrations of isotope carry-over effects in the overnight culture using ^1H , ^{13}C HSQC experiments of the metabolomes from: (a) overnight culture using ^{12}C glucose medium, post-exponential phase growth using ^{12}C glucose, (b) overnight culture using ^{12}C glucose medium, post-exponential phase growth using ^{13}C glucose, (c) overnight culture using ^{13}C glucose medium, post-exponential phase growth using ^{12}C glucose, and (d) overnight culture using ^{13}C glucose medium, post-exponential phase growth using ^{13}C glucose.

3.3.5 Storage of metabolite samples

The metabolite samples should be lyophilized and kept in -80°C freezer for long-term storage. Once the sample is prepared in NMR tubes, only short term storage in 4°C for 1-2 days is allowed, as a short term in the room temperature can lead to observable changes (Figure 3.4). This is important to notice when running multiple samples for long experiments such as ^1H , ^{13}C HSQC experiments. The samples still in the queue should be kept in 4°C refrigerator. It is also advisable to add 0.5 mM NaN_3 to inhibit trace bacterial activities.⁷⁶

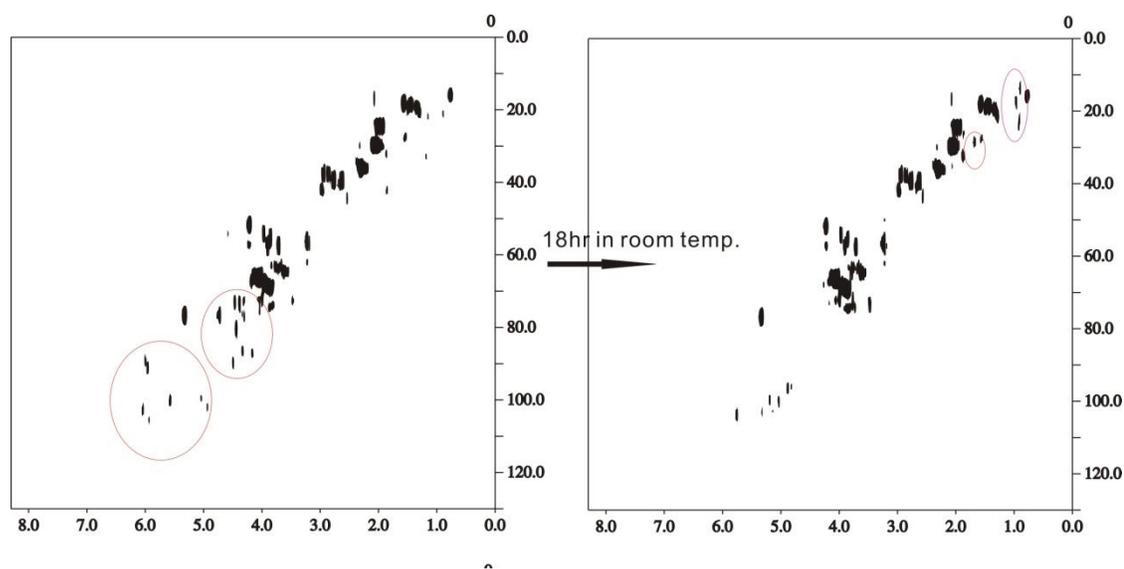


Figure 3.4 Metabolic changes dependent on the time of storage. ^1H , ^{13}C HSQC spectra of the same sample: *S. aureus* wild-type Newman strain metabolome in post-exponential phase.⁷⁷

3.4 NMR Spectroscopy

3.4.1 One-dimensional ^1H NMR methodology

One-dimensional (1D) ^1H (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 ^1H 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-d $_4$ (TMSP-d $_4$, Sigma). Thus, 1D ^1H NMR experiments combined with multivariate statistics are commonly used for the global analysis of the metabolome.

Collecting 1D ^1H 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 ^1H 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 3.5). 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 3.6). The statistical significance of cluster separation as measured by p -values is shown to decrease as both a function of group variance and the number of replicates.⁷⁸ 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 $\leq 2\text{K}$.

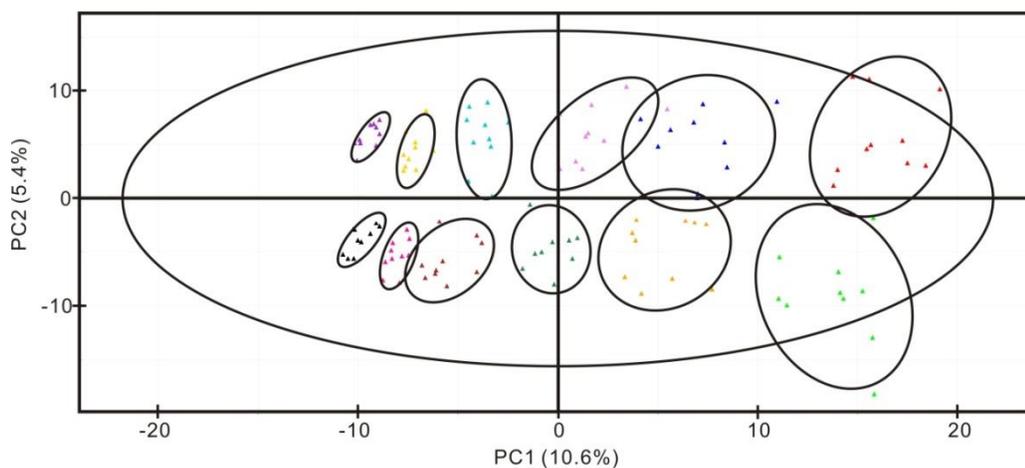


Figure 3.5 Illustration of the impact of the NMR signal-to-noise on within class variation in a PCA scores plot. From right to left, the 1D ^1H NMR spectra were collected with an increasing number of scans (1, 2, 4, 8, 16, and 32) resulting in a proportional increase in signal-to-noise. All other experimental parameters were kept constant.

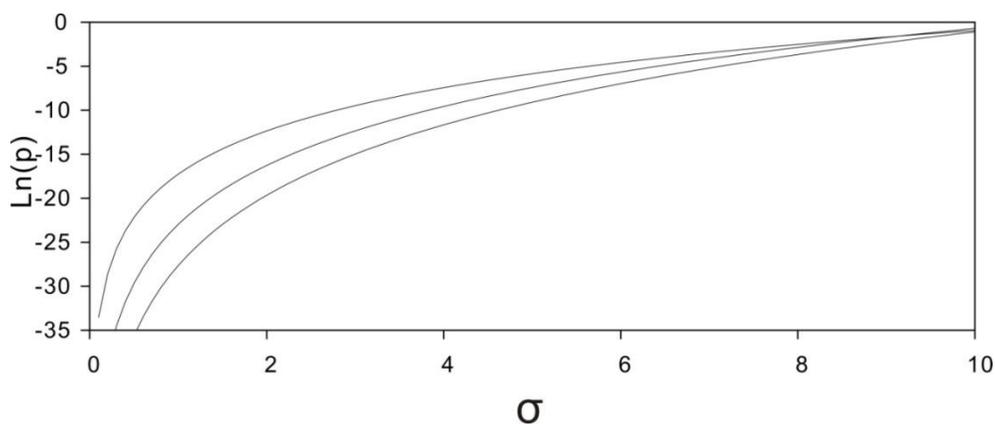


Figure 3.6 Illustration of the impact of within group variation and the number of replicates on the p values calculated between clusters in a simulated PCA scores plot. From top to bottom, p values from the simulated PCA scores plot were calculated with an increasing number of replicates (6, 8, 10) resulting 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.

A D₂O phosphate buffer is the typical solvent of choice for aqueous metabolomics samples in order to efficiently remove residual water signals and avoid interference from 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.⁷⁹ 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.⁸⁰⁻⁸⁴ 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).⁸⁴ 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.^{84,85} 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.⁸⁴ SOGGY completely eliminates the need to apply any baseline correction.

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

The severe overlap of signals in a 1D ^1H 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 ^1H NMR spectra to assign metabolites. Instead, we routinely use 2D ^1H , ^{13}C HSQC experiments for metabolite assignments. The 2D ^1H , ^{13}C HSQC experiment is a more reliable approach for metabolite identification because of the significantly higher resolution and the correlation between ^1H and ^{13}C chemical shifts for each C-H pair in a molecule.^{86,87} Also, the 2D ^1H , ^{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 ^1H , ^{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 ^1H 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 ^1H axis (~ 7.0 ppm). In general, the 2D ^1H , ^{13}C HSQC experiment requires approximately 4 hours per sample on our system.

A conventional 2D ^1H , ^{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 ^1H NMR spectra, obtaining metabolite concentrations is more difficult because peak intensities are dependent on J-couplings, dynamics and relaxation, in addition to metabolite concentrations.^{88,89} To quantify absolute metabolite concentrations, we use the Time-Zero HSQC (HSQC₀) experiment.⁸⁸ This approach requires collecting a series of three HSQC spectra (HSQC₁, HSQC₂, HSQC₃) 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 HSQC₀ 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 HSQC₂ and HSQC₃. 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 HSQC₀ set of experiment requires approximately 6 hours per sample on our system.

3.5 Data analysis

3.5.1 Preprocessing of 1D ^1H NMR data

The 1D ^1H 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.^{90,91} 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:

$$Z = \frac{X_i - \bar{X}}{\sigma} \quad (1)$$

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 3.7a). 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 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

3.7b). The protocol for identifying noise regions has been recently improved upon and results in smaller within class variations (Figure 3.7c). 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 3.1×10^{-13} to 8.1×10^{-15} , while the small, but biologically irrelevant, separation between clusters 2hwt and 2hacna (p -value 2.5×10^{-3}) was removed (Figure 3.7). 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 the average 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:

$$\text{Noise: } \sigma_i' \leq \sigma_0' \quad (2)$$

$$\text{Cutoff: } \sigma_0' = \text{avg}(\sigma_n') + 2 * \text{sd}(\sigma_n') \quad (3)$$

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 $\text{avg}(\sigma_n')$ and $\text{sd}(\sigma_n')$ 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.

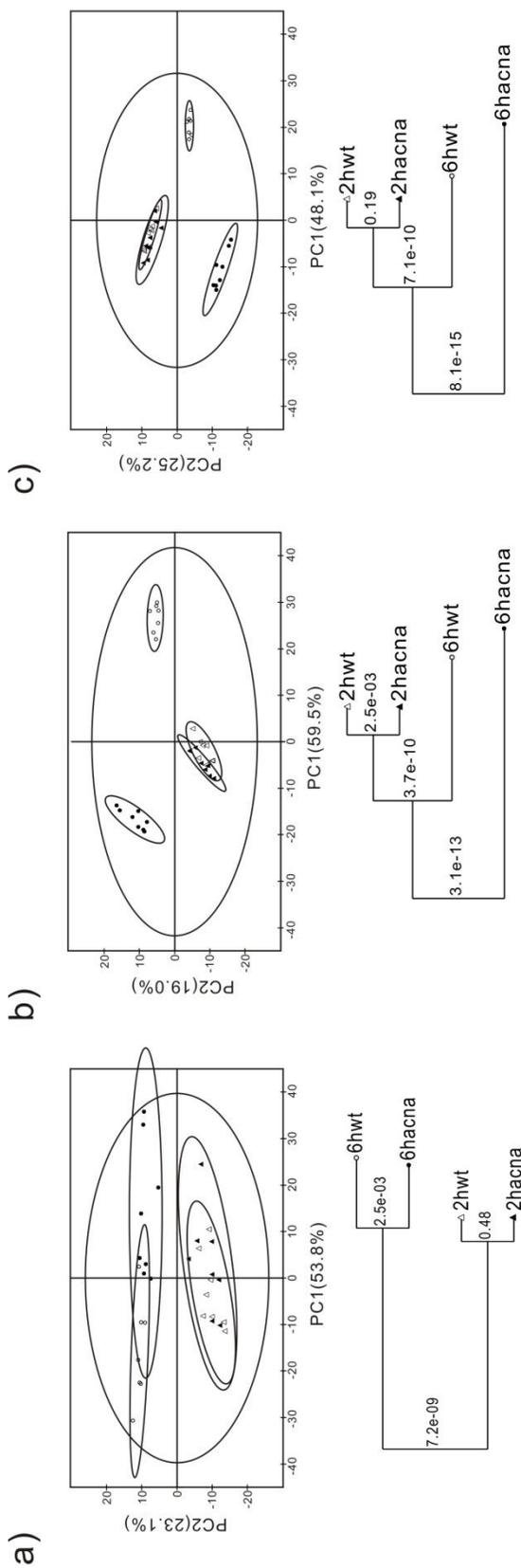


Figure 3.7 Illustration of the impact of NMR preprocessing on within and between class variations in a PCA scores plot. a) The 1D ^1H NMR spectra was not properly preprocessed. The spectra were not normalized and the noise was not removed. The spectra were only Fourier transformed, phased corrected, and the residual H_2O resonance was removed. b) The 1D ^1H NMR spectra were processed as in a) with the addition of normalization using center averaging, but without noise removal. c) The 1D ^1H 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.

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.^{90,91} In our experience, unit variance scaling, also known as auto-scaling or a Z-score (see eqn. 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 auto-scaling, which is our default scaling method.

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

We routinely apply PCA, a non-supervised technique, to determine if the 1D ^1H 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,⁹² 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^{93,94} and CV-ANOVA.⁹⁵ 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,⁹⁶ 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 ($R^2 \gg Q^2$) does suggest an over-fit model. Conversely, a p -value $\ll 0.05$ 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.⁹⁷ This finding is also supported by the increase in p -values seen with an increase in within class variations (Figure 3.6). 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 package⁷⁸ 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.^{78,97} 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,⁹⁸ 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 $\ll 0.05$ 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 ^1H chemical shifts that are correlated or anti-correlated with the separation between the two classes in an OPLS-DA scores plot. The ^1H chemical shifts can then be compared against a number of online NMR metabolomics databases to assign the metabolites.⁹⁹⁻¹⁰³ 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 ^1H , ^{13}C HSQC and ^1H , ^1H 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 ^1H chemical shifts for each J-coupled H pair; and the correlation between ^1H 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.¹⁰⁴ 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-d₄). 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 ¹H and ¹³C 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 3.8).

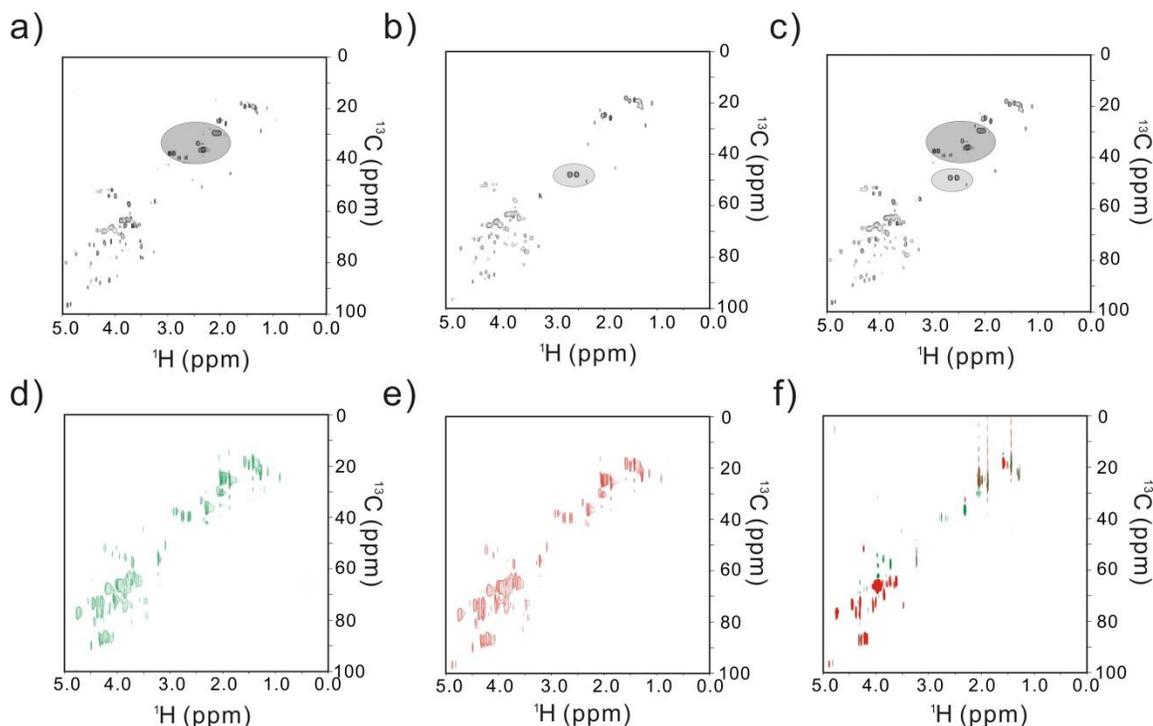


Figure 3.8 (a-c) Illustration of the procedure to generate a “master spectrum” and facilitate automated peak picking by creating a complete peak list. (a-b) Representative 2D ^1H , ^{13}C HSQC spectra obtained from two distinct bacterial cultures, where some major spectral differences are highlighted. c) The two 2D ^1H , ^{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 ^1H , ^{13}C HSQC spectra obtained from two distinct bacterial cultures. f) The two 2D ^1H , ^{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 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 metabolite spectral assignment relies largely on the completeness of metabolomics databases. We routinely use a combination of the following databases: Human Metabolome Database,⁹⁹ Madison Metabolomics Consortium Database,¹⁰⁰ Platform for RIKEN Metabolomics,¹⁰¹ BioMagResBank,¹⁰² and Metabominer,¹⁰³ 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 database 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 ^1H , ^{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 ^1H , ^{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 Biocyc¹⁰⁵ and KEGG¹⁰⁶ 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 ^1H , ^{13}C HSQC data.

After assigning the 2D ^1H , ^{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 ^1H , ^{13}C HSQC experiment are dependent on metabolite concentrations and J-couplings, dynamics, and relaxation properties.⁸⁸ Therefore, only a relative percent change in a metabolite concentration can be determined.⁵ Alternatively, an absolute concentration can be determined using HSQC₀, which requires a set of three HSQC experiments per sample. We routinely employ both approaches.⁸⁸

A relative difference in peak intensities is determined by using a triplicate set of a conventional 2D ^1H , ^{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-d₄ 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 HSQC₀ 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.^{88,89} 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 3.9 illustrates such a calibration curve using 5 different mixtures, each consisting of 9 different ¹³C-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 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 HSQC₀ 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-d₄ 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-d₄) 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.

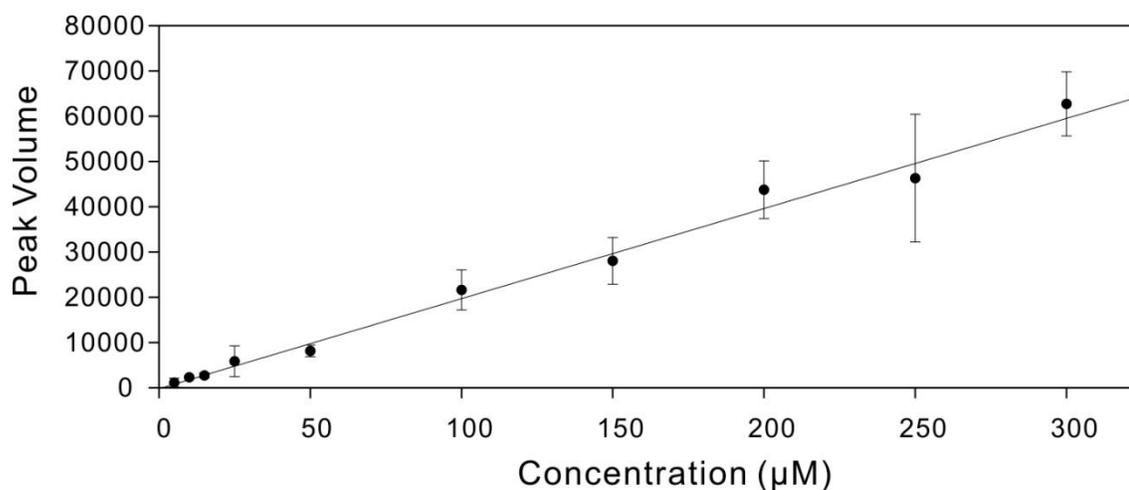


Figure 3.9 A strong correlation between NMR peak volumes and metabolite concentrations (R^2 0.997) is demonstrated by linear regression plot generated from HSQC₀ data. HSQC₀ NMR spectra were collected for five different metabolite mixtures containing nine ¹³C-labeled compounds with concentrations ranging from 5 µM to 300 µM. The relationship between peak volume and metabolite concentration is independent of the metabolite.

3.5.4 Metabolomics Network Map

Metabolites are highly interconnected through numerous metabolic pathways that form an extremely complex network.¹⁰⁷ 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.¹⁰⁸⁻¹¹¹ 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.¹¹² 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.¹¹³ 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 misassigned 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. The structure is made of nodes and edges. Metabolites are small molecules that are involved in the metabolism which is the overall chemical reactions in the biological system. Therefore, in terms of choosing metabolites,

the intermediates from glycolysis pathways, TCA cycle and pentose phosphate pathways were taken as the major contributions since they are essential to synthesis all the macromolecules for *S. aureus*, although for pentose phosphate pathway, only a portion of it was selected because there are many metabolites that are not detectable by our technique so far. Other metabolites, especially the common amino acids that can be derived from these three pathways directly were also included. The goal is to maximize the efficiency of the usage of the network so that it should represent the status of the cells and provide systematic information on the perturbed cellular metabolome. For each project, it can be modified as needed. The connections between different nodes are based on *S. aureus* pathway stored in the *MetaCyc* database. The direct and indirect reactions were all represented by edges without discrimination because multiple pathways can happen. Therefore only the shortest one was shown. If a metabolite increases its concentration, red color is assigned. If a metabolite decreases its color, green color is assigned. If there is no change between the comparison pair, then a grey color is assigned. The size of the node representing the metabolite is increased or decreased in proportion to the relative change. For example, if metabolite A is 2 times increased in its concentration, the node is 2 times larger than the ones with no change and vice versa. The exact number of times is normalized from 1 to 2 in order to keep the basic shape of the map, without being perturbed by some huge nodes that may be made. A student's t-test is used to determine that if a concentration is significantly changed or not, only p value smaller than 0.05 are taken to be changed in concentration. Under other conditions, if a change has a p value larger than 0.05 but smaller than 0.1, a pink or light green was coded. The layout

was designed based on traditional flowchart. The glycolysis and TCA cycle are highlighted by broadening the width of the edges.

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.; Yang, N. S. *Current opinion in chemical biology* **2008**, *12*, 66.
- (2) Spratlin, J. L.; Serkova, N. J.; Eckhardt, S. G. *Clin Cancer Res* **2009**, *15*, 431.
- (3) Rochfort, S. *Journal of natural products* **2005**, *68*, 1813.
- (4) Raamsdonk, L. M.; Teusink, B.; Broadhurst, D.; Zhang, N.; Hayes, A.; Walsh, M. C.; Berden, J. A.; Brindle, K. M.; Kell, D. B.; Rowland, J. J.; Westerhoff, H. V.; van Dam, K.; Oliver, S. G. *Nature biotechnology* **2001**, *19*, 45.
- (5) Zhang, B.; Halouska, S.; Schiaffo, C. E.; Sadykov, M. R.; Somerville, G. A.; Powers, R. *J Proteome Res* **2011**, *10*, 3743.
- (6) Ewald, J. C.; Matt, T.; Zamboni, N. *Molecular bioSystems* **2013**, *9*, 440.
- (7) Fiehn, O. *Plant molecular biology* **2002**, *48*, 155.
- (8) Kaddurah-Daouk, R.; Kristal, B. S.; Weinshilboum, R. M. *Annual review of pharmacology and toxicology* **2008**, *48*, 653.
- (9) Keun, H. C. *Pharmacology & therapeutics* **2006**, *109*, 92.
- (10) Sadykov, M. R.; Zhang, B.; Halouska, S.; Nelson, J. L.; Kreimer, L. W.; Zhu, Y.; Powers, R.; Somerville, G. A. *J Biol Chem* **2010**, *285*, 36616.
- (11) Ideker, T.; Thorsson, V.; Ranish, J. A.; Christmas, R.; Buhler, J.; Eng, J. K.; Bumgarner, R.; Goodlett, D. R.; Aebersold, R.; Hood, L. *Science* **2001**, *292*, 929.
- (12) Pir, P.; Kirdar, B.; Hayes, A.; Onsan, Z. Y.; Ulgen, K. O.; Oliver, S. G. *BMC bioinformatics* **2006**, *7*, 203.

- (13) Mosier, A. C.; Justice, N. B.; Bowen, B. P.; Baran, R.; Thomas, B. C.; Northen, T. R.; Banfield, J. F. *mBio* **2013**, *4*.
- (14) Sadykov, M. R.; Zhang, B.; Halouska, S.; Nelson, J. L.; Kreimer, L. W.; Zhu, Y. F.; Powers, R.; Somerville, G. A. *Journal of Biological Chemistry* **2010**, *285*, 36616.
- (15) Yoon, S. H.; Han, M. J.; Jeong, H.; Lee, C. H.; Xia, X. X.; Lee, D. H.; Shim, J. H.; Lee, S. Y.; Oh, T. K.; Kim, J. F. *Genome biology* **2012**, *13*, R37.
- (16) Yang, R.; Du, Z.; Han, Y.; Zhou, L.; Song, Y.; Zhou, D.; Cui, Y. *Frontiers in cellular and infection microbiology* **2012**, *2*, 157.
- (17) Bisson, G. P.; Mehaffy, C.; Broeckling, C.; Prenni, J.; Rifat, D.; Lun, D. S.; Burgos, M.; Weissman, D.; Karakousis, P. C.; Dobos, K. *Journal of bacteriology* **2012**, *194*, 6441.
- (18) Nobeli, I.; Ponstingl, H.; Krissinel, E. B.; Thornton, J. M. *J. Mol. Biol.* **2003**, *334*, 697.
- (19) Weckwerth, W. *Bioanalysis* **2010**, *2*, 829.
- (20) Powers, R. *Magn Reson Chem* **2009**, *47 Suppl 1*, S2.
- (21) Moco, S.; Bino, R. J.; De Vos, R. C. H.; Vervoort, J. *Trac-Trends in Analytical Chemistry* **2007**, *26*, 855.
- (22) Villas-Boas, S. G.; Mas, S.; Aakesson, M.; Smedsgaard, J.; Nielsen, J. *Mass Spectrom. Rev.* **2005**, *24*, 613.
- (23) Wilson, I. D.; Plumb, R.; Granger, J.; Major, H.; Williams, R.; Lenz, E. *M. J. Chromatogr., B: Anal. Technol. Biomed. Life Sci.* **2005**, *817*, 67.
- (24) Reo, N. V. *Drug Chem. Toxicol.* **2002**, *25*, 375.
- (25) Fan, T. W. M.; Lane, A. N. *J. Biomol. NMR* **2011**, *49*, 267.

- (26) Lewis, I. A.; Karsten, R. H.; Norton, M. E.; Tonelli, M.; Westler, W. M.; Markley, J. L. *Anal. Chem. (Washington, DC, U. S.)* **2010**, *82*, 4558.
- (27) Kell, D. B. *Curr. Opin. Microbiol.* **2004**, *7*, 296.
- (28) Metz, T. O.; Page, J. S.; Baker, E. S.; Tang, K.; Ding, J.; Shen, Y.; Smith, R. D. *TrAC, Trends Anal. Chem.* **2008**, *27*, 205.
- (29) Pan, Z.; Raftery, D. *Anal. Bioanal. Chem.* **2007**, *387*, 525.
- (30) Viant, M. R.; Rosenblum, E. S.; Tjeerdema, R. S. *Environmental Science & Technology* **2003**, *37*, 4982.
- (31) Beckonert, O.; Keun, H. C.; Ebbels, T. M. D.; Bundy, J.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. *Nat. Protoc.* **2007**, *2*, 2692.
- (32) Dietmair, S.; Timmins, N. E.; Gray, P. P.; Nielsen, L. K.; Krömer, J. O. *Analytical Biochemistry* **2010**, *404*, 155.
- (33) Kim, H. K.; Choi, Y. H.; Verpoorte, R. *Nature protocols* **2010**, *5*, 536.
- (34) Marcinowska, R.; Trygg, J.; Wolf-Watz, H.; Mortiz, T.; Surowiec, I. *J Microbiol Methods* **2011**, *87*, 24.
- (35) Mashego, M.; Rumbold, K.; De Mey, M.; Vandamme, E.; Soetaert, W.; Heijnen, J. *Biotechnology Letters* **2007**, *29*, 1.
- (36) Wu, X. H.; Yu, H. L.; Ba, Z. Y.; Chen, J. Y.; Sun, H. G.; Han, B. Z. *Biotechnol J* **2010**, *5*, 75.
- (37) Cuperlovic-Culf, M.; Barnett, D. A.; Culf, A. S.; Chute, I. *Drug Discovery Today* **2010**, *15*, 610.
- (38) Ramadan, Z.; Jacobs, D.; Grigorov, M.; Kochhar, S. *Talanta* **2006**, *68*, 1683.

- (39) Bylesjo, M.; Rantalainen, M.; Cloarec, O.; Nicholson, J. K.; Holmes, E.; Trygg, J. *Journal of Chemometrics* **2006**, *20*, 341.
- (40) Fernie, A. R.; Trethewey, R. N.; Krotzky, A. J.; Willmitzer, L. *Nature Reviews Molecular Cell Biology* **2004**, *5*, 763.
- (41) Clarke, C. J.; Haselden, J. N. *Toxicologic Pathology* **2008**, *36*, 140.
- (42) Faijes, M.; Mars, A. E.; Smid, E. J. *Microbial Cell Factories* **2007**, *6*, 27.
- (43) Van Batenburg, M. F.; Coulier, L.; van Eeuwijk, F.; Smilde, A. K.; Westerhuis, J. A. *Anal. Chem. (Washington, DC, U. S.)* **2011**, *83*, 3267.
- (44) Kanani, H.; Chrysanthopoulos, P. K.; Klapa, M. I. *J. Chromatogr. B* **2008**, *871*, 191.
- (45) Teahan, O.; Gamble, S.; Holmes, E.; Waxman, J.; Nicholson, J. K.; Bevan, C.; Keun, H. C. *Anal. Chem.* **2006**, *78*, 4307.
- (46) Worley, B.; Powers, R. *Current Metabolomics* **2013**, *1*, 92.
- (47) Birkemeyer, C.; Luedemann, A.; Wagner, C.; Erban, A.; Kopka, J. *Trends Biotechnol* **2005**, *23*, 28.
- (48) Fan, T. M.; Bandura, L.; Higashi, R.; Lane, A. *Metabolomics* **2005**, *1*, 325.
- (49) Halouska, S.; Fenton, R. J.; Barletta, R. G.; Powers, R. *ACS chemical biology* **2012**, *7*, 166.
- (50) Halouska, S.; Chacon, O.; Fenton, R. J.; Zinniel, D. K.; Barletta, R. G.; Powers, R. *J Proteome Res* **2007**, *6*, 4608.
- (51) Forgue, P.; Halouska, S.; Werth, M.; Xu, K. M.; Harris, S.; Powers, R. *Journal of Proteome Research* **2006**, *5*, 1916.

- (52) Suhre, K.; Gieger, C. *Nature Reviews Genetics* **2012**, *13*, 759.
- (53) Halouska, S.; Powers, R. *Journal of magnetic resonance* **2006**, *178*, 88.
- (54) Bolten, C. J.; Kiefer, P.; Letisse, F.; Portais, J. C.; Wittmann, C. *Analytical chemistry* **2007**, *79*, 3843.
- (55) Bailey, N. J. C.; Oven, M.; Holmes, E.; Zenk, M. H.; Nicholson, J. K. *Spectroscopy-an International Journal* **2004**, *18*, 279.
- (56) Defernez, M.; Colquhoun, I. J. *Phytochemistry* **2003**, *62*, 1009.
- (57) Broadhurst, D. I.; Kell, D. B. *Metabolomics* **2006**, *2*, 171.
- (58) Brereton, R. G. *Trac-Trends in Analytical Chemistry* **2006**, *25*, 1103.
- (59) Canelas, A.; Ras, C.; ten Pierick, A.; van Dam, J.; Heijnen, J.; van Gulik, W. *Metabolomics* **2008**, *4*, 226.
- (60) Saude, E. J.; Sykes, B. D. *Metabolomics* **2007**, *3*, 19.
- (61) Giuliadori, A. M.; Gualerzi, C. O.; Soto, S.; Vila, J.; Tavio, M. M. *Ann. N. Y. Acad. Sci.* **2007**, *1113*, 95.
- (62) Schadel, F.; David, F.; Franco-Lara, E. *Applied Microbiology and Biotechnology* **2011**, *92*, 1261.
- (63) Wellerdiek, M.; Winterhoff, D.; Reule, W.; Brandner, J.; Oldiges, M. *Bioprocess and Biosystems Engineering* **2009**, *32*, 581.
- (64) Wittmann, C.; Kromer, J. O.; Kiefer, P.; Binz, T.; Heinzle, E. *Analytical Biochemistry* **2004**, *327*, 135.
- (65) Mazo, R. M. *J. Phys. Chem. B* **2006**, *110*, 24077.

- (66) Gardner, K. H.; Kay, L. E. *Annu Rev Biophys Biomol Struct* **1998**, *27*, 357.
- (67) Lundstrom, P.; Teilum, K.; Carstensen, T.; Bezsonova, I.; Wiesner, S.; Hansen, D. F.; Religa, T. L.; Akke, M.; Kay, L. E. *J Biomol NMR* **2007**, *38*, 199.
- (68) Schoenheimer, R.; Rittenberg, D. *Science* **1935**, *82*, 156.
- (69) Wasylenko, T. M.; Stephanopoulos, G. *Biotechnol J* **2013**, *8*, 1080.
- (70) Kikuchi, J.; Shinozaki, K.; Hirayama, T. *Plant Cell Physiol* **2004**, *45*, 1099.
- (71) Kikuchi, J.; Hirayama, T. *Methods Mol Biol* **2007**, *358*, 273.
- (72) Fan, T.; Lane, A.; Higashi, R.; Farag, M.; Gao, H.; Bousamra, M.; Miller, D. *Molecular Cancer* **2009**, *8*, 41.
- (73) Gowda, G. A.; Shanaiah, N.; Raftery, D. *Adv Exp Med Biol* **2012**, *992*, 147.
- (74) Nuxoll, A. S.; Halouska, S. M.; Sadykov, M. R.; Hanke, M. L.; Bayles, K. W.; Kielian, T.; Powers, R.; Fey, P. D. *PLoS Pathog* **2012**, *8*, e1003033.
- (75) McKee, A. S.; McDermid, A. S.; Ellwood, D. C.; Marsh, P. D. *J Appl Bacteriol* **1985**, *59*, 263.
- (76) Lewis, I. A.; Schommer, S. C.; Hodis, B.; Robb, K. A.; Tonelli, M.; Westler, W. M.; Sussman, M. R.; Markley, J. L. *Anal Chem* **2007**, *79*, 9385.
- (77) Hartmann, T.; Zhang, B.; Baronian, G.; Schulthess, B.; Homerova, D.; Grubmuller, S.; Kutzner, E.; Gaupp, R.; Bertram, R.; Powers, R.; Eisenreich, W.; Kormanec, J.; Herrmann, M.; Molle, V.; Somerville, G. A.; Bischoff, M. *J Biol Chem* **2013**, *288*, 36116.
- (78) Worley, B.; Halouska, S.; Powers, R. *Anal. Biochem.* **2013**, *433*, 102.

- (79) Potts, B. C. M.; Deese, A. J.; Stevens, G. J.; Reily, M. D.; Robertson, D. G.; Theiss, J. J. *Pharm. Biomed. Anal.* **2001**, *26*, 463.
- (80) Hoult, D. I. *Journal of magnetic resonance* **1976**, *21*, 337.
- (81) Sklenar, V.; Piotto, M.; Leppik, R.; Saudek, V. *Journal of Magnetic Resonance Series A* **1993**, *102*, 241.
- (82) Liu, M. L.; Mao, X. A.; Ye, C. H.; Huang, H.; Nicholson, J. K.; Lindon, J. C. *Journal of magnetic resonance* **1998**, *132*, 125.
- (83) Ogg, R. J.; Kingsley, P. B.; Taylor, J. S. *Journal of Magnetic Resonance Series B* **1994**, *104*, 1.
- (84) Simpson, A. J.; Brown, S. A. *Journal of magnetic resonance* **2005**, *175*, 340.
- (85) Hwang, T. L.; Shaka, A. J. *Journal of Magnetic Resonance Series A* **1995**, *112*, 275.
- (86) McKenzie, J.; Charlton, A.; Donarski, J.; MacNicoll, A.; Wilson, J. *Metabolomics* **2010**, *6*, 574.
- (87) Xi, Y.; de Ropp, J. S.; Viant, M. R.; Woodruff, D. L.; Yu, P. *Analytica chimica acta* **2008**, *614*, 127.
- (88) Hu, K. F.; Westler, W. M.; Markley, J. L. *Journal of the American Chemical Society* **2011**, *133*, 1662.
- (89) Hu, K. F.; Ellinger, J. J.; Chylla, R. A.; Markley, J. L. *Analytical Chemistry* **2011**, *83*, 9352.
- (90) van den Berg, R. A.; Hoefsloot, H. C.; Westerhuis, J. A.; Smilde, A. K.; van der Werf, M. J. *BMC genomics* **2006**, *7*, 142.

- (91) Craig, A.; Cloareo, O.; Holmes, E.; Nicholson, J. K.; Lindon, J. C. *Analytical Chemistry* **2006**, 78, 2262.
- (92) Kjeldahl, K.; Bro, R. *Journal of Chemometrics* **2010**, 24, 558.
- (93) Shao, J. *Journal of the American Statistical Association* **1993**, 88, 486.
- (94) Golbraikh, A.; Tropsha, A. *Journal of Molecular Graphics & Modelling* **2002**, 20, 269.
- (95) Eriksson, L.; Trygg, J.; Wold, S. *Journal of Chemometrics* **2008**, 22, 594.
- (96) Westerhuis, J.; Hoefsloot, H.; Smit, S.; Vis, D.; Smilde, A.; van Velzen, E.; van Duijnhoven, J.; van Dorsten, F. *Metabolomics* **2008**, 4, 81.
- (97) Werth, M. T.; Halouska, S.; Shortridge, M. D.; Zhang, B.; Powers, R. *Analytical Biochemistry* **2010**, 399, 58.
- (98) Henderson, A. R. *Clin. Chim. Acta* **2005**, 359, 1.
- (99) Wishart, D. S.; Knox, C.; Guo, A. C.; Eisner, R.; Young, N.; Gautam, B.; Hau, D. D.; Psychogios, N.; Dong, E.; Bouatra, S.; Mandal, R.; Sinelnikov, I.; Xia, J. G.; Jia, L.; Cruz, J. A.; Lim, E.; Sobsey, C. A.; Shrivastava, S.; Huang, P.; Liu, P.; Fang, L.; Peng, J.; Fradette, R.; Cheng, D.; Tzur, D.; Clements, M.; Lewis, A.; De Souza, A.; Zuniga, A.; Dawe, M.; Xiong, Y. P.; Clive, D.; Greiner, R.; Nazyrova, A.; Shaykhtudinov, R.; Li, L.; Vogel, H. J.; Forsythe, I. *Nucleic Acids Research* **2009**, 37, D603.
- (100) Cui, Q.; Lewis, I. A.; Hegeman, A. D.; Anderson, M. E.; Li, J.; Schulte, C. F.; Westler, W. M.; Eghbalnia, H. R.; Sussman, M. R.; Markley, J. L. *Nature biotechnology* **2008**, 26, 162.
- (101) Akiyama, K.; Chikayama, E.; Yuasa, H.; Shimada, Y.; Tohge, T.; Shinozaki, K.; Hirai, M. Y.; Sakurai, T.; Kikuchi, J.; Saito, K. *In silico biology* **2008**, 8, 339.

- (102) Ulrich, E. L.; Akutsu, H.; Doreleijers, J. F.; Harano, Y.; Ioannidis, Y. E.; Lin, J.; Livny, M.; Mading, S.; Maziuk, D.; Miller, Z.; Nakatani, E.; Schulte, C. F.; Tolmie, D. E.; Kent Wenger, R.; Yao, H.; Markley, J. L. *Nucleic acids research* **2008**, *36*, D402.
- (103) Xia, J.; Bjorndahl, T. C.; Tang, P.; Wishart, D. S. *BMC bioinformatics* **2008**, *9*, 507.
- (104) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *Journal of biomolecular NMR* **1995**, *6*, 277.
- (105) Caspi, R.; Altman, T.; Dreher, K.; Fulcher, C. A.; Subhraveti, P.; Keseler, I. M.; Kothari, A.; Krummenacker, M.; Latendresse, M.; Mueller, L. A.; Ong, Q.; Paley, S.; Pujar, A.; Shearer, A. G.; Travers, M.; Weerasinghe, D.; Zhang, P.; Karp, P. D. *Nucleic acids research* **2012**, *40*, D742.
- (106) Kanehisa, M.; Araki, M.; Goto, S.; Hattori, M.; Hirakawa, M.; Itoh, M.; Katayama, T.; Kawashima, S.; Okuda, S.; Tokimatsu, T.; Yamanishi, Y. *Nucleic Acids Res.* **2008**, *36*, D480.
- (107) Kohlstedt, M.; Becker, J.; Wittmann, C. *Applied Microbiology and Biotechnology* **2010**, *88*, 1065.
- (108) Bot, J. J.; Reinders, M. J. *Bioinformatics* **2011**, *27*, 2451.
- (109) Gao, J.; Tarcea, V. G.; Karnovsky, A.; Mirel, B. R.; Weymouth, T. E.; Beecher, C. W.; Cavalcoli, J. D.; Athey, B. D.; Omenn, G. S.; Burant, C. F.; Jagadish, H. V. *Bioinformatics* **2010**, *26*, 971.
- (110) Kohl, M.; Wiese, S.; Warscheid, B. *Methods in molecular biology* **2011**, *696*, 291.
- (111) Smoot, M. E.; Ono, K.; Ruscheinski, J.; Wang, P. L.; Ideker, T. *Bioinformatics* **2011**, *27*, 431.

(112) Caspi, R.; Foerster, H.; Fulcher, C. A.; Kaipa, P.; Krummenacker, M.; Latendresse, M.; Paley, S.; Rhee, S. Y.; Shearer, A. G.; Tissier, C.; Walk, T. C.; Zhang, P.; Karp, P. D. *Nucleic Acids Res.* **2008**, *36*, D623.

(113) Chaika, N. V.; Gebregiworgis, T.; Lewallen, M. E.; Purohit, V.; Radhakrishnan, P.; Liu, X.; Zhang, B.; Mehla, K.; Brown, R. B.; Caffrey, T.; Yu, F.; Johnson, K. R.; Powers, R.; Hollingsworth, M. A.; Singh, P. K. *Proc Natl Acad Sci U S A* **2012**, *109*, 13787.

CHAPTER 4

AN INEXPENSIVE HIGH-THROUGHPUT NMR TUBE CLEANING APPARATUS

4.1 Introduction

High-throughput screening (HTS) is widely used in the biotechnology and pharmaceutical industries¹⁻³ with an expanding interest in academia.^{4,5} HTS is an efficient approach for the experimental testing of large sample sets for the discovery of new drugs. Multiple reviews on HTS automation have discussed robotic systems employed to transport samples, add reagents, mix samples or detect signals.⁶⁻⁸ These steps are clearly essential components to a successful screen and are understandably a focal-point of any assay design. Conversely, there is a complete lack of reports discussing the automation of recycling devices. For the majority of high-throughput screens, a reliance on disposable labware (assay plates, pipette tips, *etc*) simplifies the execution of the assay while being reasonably cost-effective.

Fragment-based NMR screens complement traditional HTS and have evolved to become an important and common component of drug discovery.^{9,10} Fragment-based screens utilize a chemical library of low molecule-weight ligands ($\leq 200 - 300$ Da)¹¹ that correspond to fragments of known drugs or have drug-like characteristics.¹² NMR is used to identify binders, to identify the ligand binding site, and to assist in “growing” the fragments to improve binding affinity. There are several advantages to fragment-based NMR screens over HTS that include maximizing ligand efficiency, improving coverage of chemical space, higher hit rates and higher quality of leads, direct observation of biologically relevant interactions, and a universal assay design. Conversely, NMR ligand

affinity screens are typically limited to assaying only hundreds to thousands of compounds. Another practical disadvantage is the hundreds of NMR tubes that quickly accumulate and require cleaning.

High quality NMR tubes are generally used for ligand affinity screens and because of cost (\$2 to \$8 per tube) are typically not limited to a single use. A standard NMR tube cleaner (Sigma-Aldrich, St. Louis, MO) can handle from one to five tubes. The NMR tube sits inverted on a glass adapter for washing, which is very fragile since it has a small diameter (<5 mm) and long length (8 in.). Thus, the routine cleaning of thousands of NMR tubes rapidly results in broken NMR tube cleaners at a cost of \$90 to \$420 per cleaner. Besides manually cleaning, there are commercial devices available such as the Bruker BioSpin Autoclean system (Billerica, MA) that is capable of cleaning 60 tubes per hour, but generates a significant amount of waste and has a cost of \$15K. Alternatively, NMR tubes can be completely discarded from the NMR screen by using a flow-probe.¹³ The NMR probe contains a fixed cell with active volumes ranging from 30 to 120 μ l. The NMR samples are sequentially transferred to the probe using a robotic liquid handling system. After data acquisition, the sample is either transferred to waste or collected. Challenges with flow probes may include cross-contamination between samples, clogging of the system by solid particulates that may also result in lost samples, and problems with properly positioning the sample in the probe. There is also the initial cost of \$75K for a flow-probe system.

A cost effective and efficient means of cleaning thousands of NMR tubes is an annoying problem and a routine challenge currently faced by NMR screening groups. To

solve this problem, an inexpensive (~\$300) and highly effective apparatus for cleaning NMR tubes has been designed and successfully employed in our laboratory. It is extremely efficient compared to existing tube cleaners since 180 tubes are cleaned in an hour. Obviously, the use of large amounts of solvent increases the cost of cleaning and diminishes the value of recycling NMR tubes. There are also safety and health concerns. As a result, the design of the NMR tube cleaner apparatus and the development of the protocol focused on minimizing solvent usage. Only 200 to 300 mL of solvent is used to clean the 180 NMR tubes. The amount of solvent waste will increase proportionally if multiple solvents are used. Similarly, an apparatus that routinely cracks, chips or scratches the NMR tubes defeats the cost-effectiveness of recycling. More concerning is the possibility of unnoticed tube damage that may result in a broken tube in an NMR probe. Unlike the manual cleaning of a large batch of NMR tubes where damage is common, we have not observed any damage to the hundreds of NMR tubes cleaned using our apparatus. This is primarily a result of minimal handling of the NMR tubes during cleaning and the fact that the tubes are simply placed inverted into a glass vessel.

4.2 Methods and Materials

Diagrams and pictures of our NMR tube cleaning apparatus are shown in Figure 4.1. The apparatus can be easily constructed in any laboratory with access to a glass blower. The cleaner consists of three major components: (i) a specially designed glass valve assembly (switch), (ii) a 750 mL Labconco® vacuum bottle, cap and adaptor, (iii) and a standard pump system (not shown). The NMR tubes are simply placed inverted into

the 750 mL Labconco® vacuum bottle. The cap is placed onto the vacuum bottle, which is then partially filled with solvent. Our glass switch is attached to the cap and the apparatus is connected to a vacuum pump. A flexibly attached glass tube extends from the glass switch to the base of the vacuum bottle. A pressure differential is used to push solvent into and out of the NMR tubes or to completely remove the solvent from the apparatus. This occurs by manipulating the two Teflon valves on our glass switch while the apparatus is attached to a vacuum pump. Basically, with the left valve closed, slowly and briefly opening the T-valve to the vacuum position evacuates the air from the Labconco® vacuum bottle and places it under a slight vacuum without removing the solvent. Turning the T-valve to the vent position opens the vacuum bottle to atmosphere breaking the vacuum and forcing solvent into the NMR tubes. Switching the T-valve back to the vacuum position places the NMR tubes back under vacuum and removes the solvent from the NMR tubes. Repeating the process two or three times effectively washes the NMR tubes with solvent. Switching the T-valve to the vent position and opening the left valve rapidly removes the solvent from the NMR tube cleaner into the filter flask. The process can be repeated with other solvents.

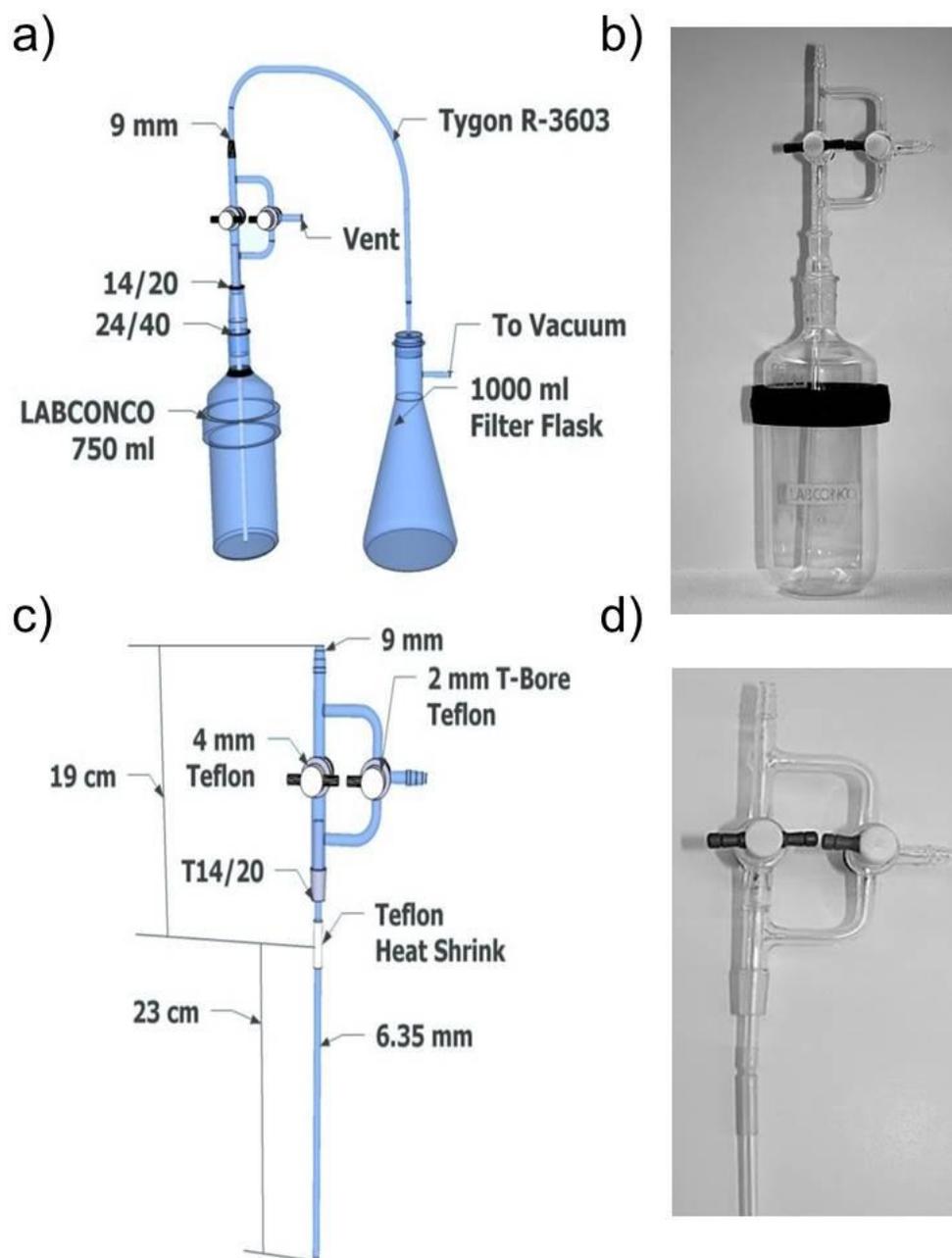


Figure 4.1 Schematic drawing a) and picture b) of our NMR tube cleaning apparatus. Schematic drawing c) and picture d) of an expanded view of the glass switch. Important features and dimensions are labeled.

The general protocol for cleaning NMR tubes (Figure 4.2) corresponds to the following steps: (1) empty the NMR tubes that need to be cleaned and remove any labels, (2) load tubes into the NMR tube cleaner along with solvent, (3) install the valve assembly and hook up the filter flask and vacuum system, (4) rinse multiple times with solvents of choice, (5) remove tubes and dry in an oven at 100 °C for an hour. For heavily soiled NMR tubes that contain solid films or stains, a simple solvent wash may not be sufficient. Instead, the NMR tubes are typically soaked overnight in a concentrated (<70%) nitric acid bath. The tubes are then quickly rinsed with tap water before placing in the NMR tube washer. The wash solvent is a saturated sodium bicarbonate solution in deionized water. Various solvents can be used in the NMR tube cleaner depending on need and the chemical composition of the original NMR samples.

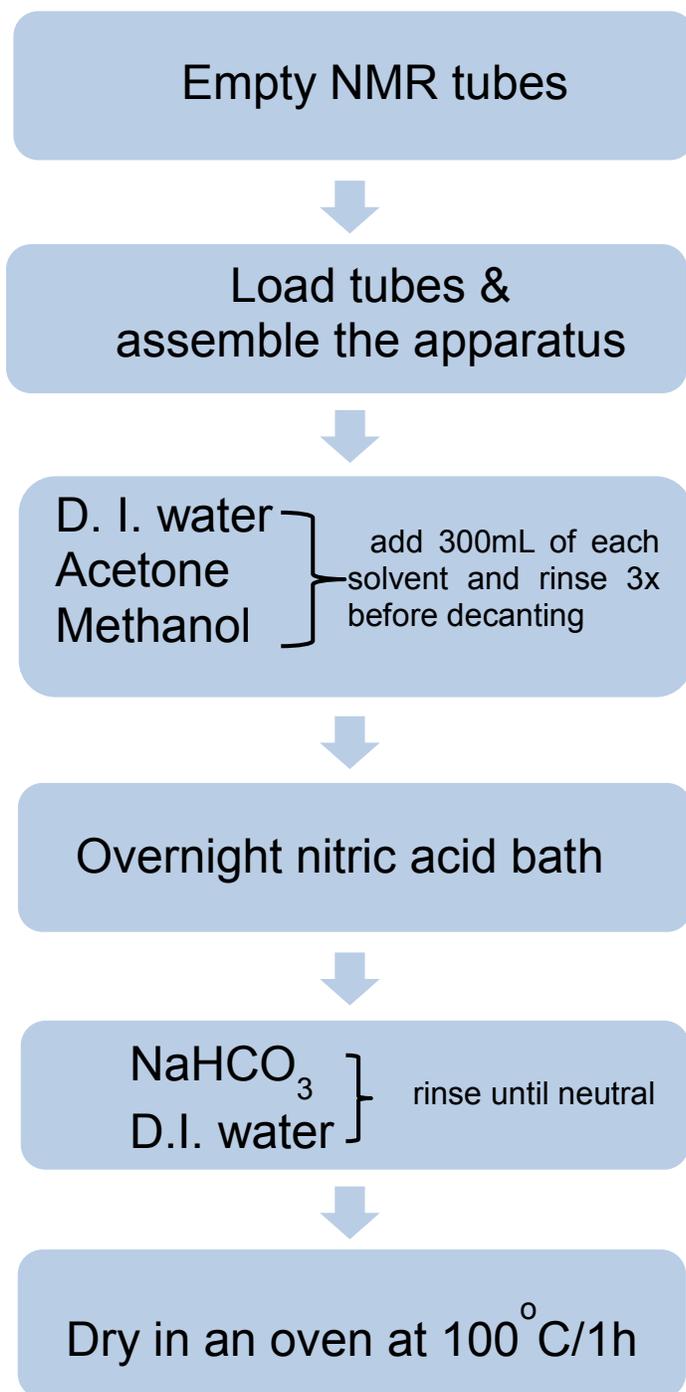


Figure 4.2 Flow diagram of the NMR tube cleaning process.

4.3 Conclusion

For a typical fragment-based NMR screen, deionized water is the first wash solvent since it will effectively remove a majority of the aqueous-based samples. A second wash using a polar organic solvent such as methanol, tetrahydrofuran or dichloromethane may also be used to remove any compound residue and promote tube drying. If necessary, a third and final wash with a non-polar solvent such as chloroform could be employed. To ensure the effectiveness of our NMR tube cleaner, there are a few general rules that need to be followed. It is important to avoid having the wash solvent sit in the NMR tubes for an extended period of time. The rapid insertion and removal of the wash solvent provides efficient agitation that assists in the cleaning process. Conversely, prolonged exposure to a wash solvent may result in compound precipitation and the formation of a difficult to remove film. For a similar reason, it is highly undesirable to allow an NMR sample to evaporate to dryness. Effectively, NMR tubes should be cleaned immediately after use to obtain optimal performance from our NMR tube cleaner.

4.1.4 References

- (1) Wu, G.; Doberstein, S. K. *Drug Discovery Today* **2006**, *11*, 718.
- (2) Kale, D. L.; Chaturvedi, S. C.; Patil, D.; Kakde, R. B. *Pharma Rev.* **2008**, *6*, 46.
- (3) Mayr, L. M.; Fuerst, P. J. *Biomol. Screening* **2008**, *13*, 443.

- (4) Frearson, J. A.; Collie, I. T. *Drug Discov Today* **2009**, *14*, 1150.
- (5) Stein, R. L. *J. Biomol. Screening* **2003**, *8*, 615.
- (6) Zhang, L.; Banks, M. N.; Houston, J. G. *Drugs Pharm. Sci.* **2009**, *196*, 6.
- (7) Hamilton, S. *Methods Mol. Biol. (Totowa, NJ, U. S.)* **2002**, *190*, 169.
- (8) Trinkka, R. F.; Leichtfried, F. E. *Drugs Pharm. Sci.* **2001**, *114*, 493.
- (9) Hajduk, P. J.; Greer, J. *Nature Reviews Drug Discovery* **2007**, *6*, 211.
- (10) Powers, R. *Expert Opin. Drug Discovery* **2009**, *4*, 1077.
- (11) Hajduk, P. J.; Galloway, W. R. J. D.; Spring, D. R. *Nature (London, U. K.)* **2011**, *470*, 42.
- (12) Lipinski, C. A. *Drug Discovery Today: Technol.* **2004**, *1*, 337.
- (13) Haner, R. L.; Llanos, W.; Mueller, L. *J. Magn. Reson.* **2000**, *143*, 69.

CHAPTER 5

NMR ANALYSIS OF A STRESS RESPONSE METABOLIC SIGNALING NETWORK

5.1 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.

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 5.1). 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 study²³ with 4% ethanol,²⁴ and Fe limitation.³⁰⁻³²

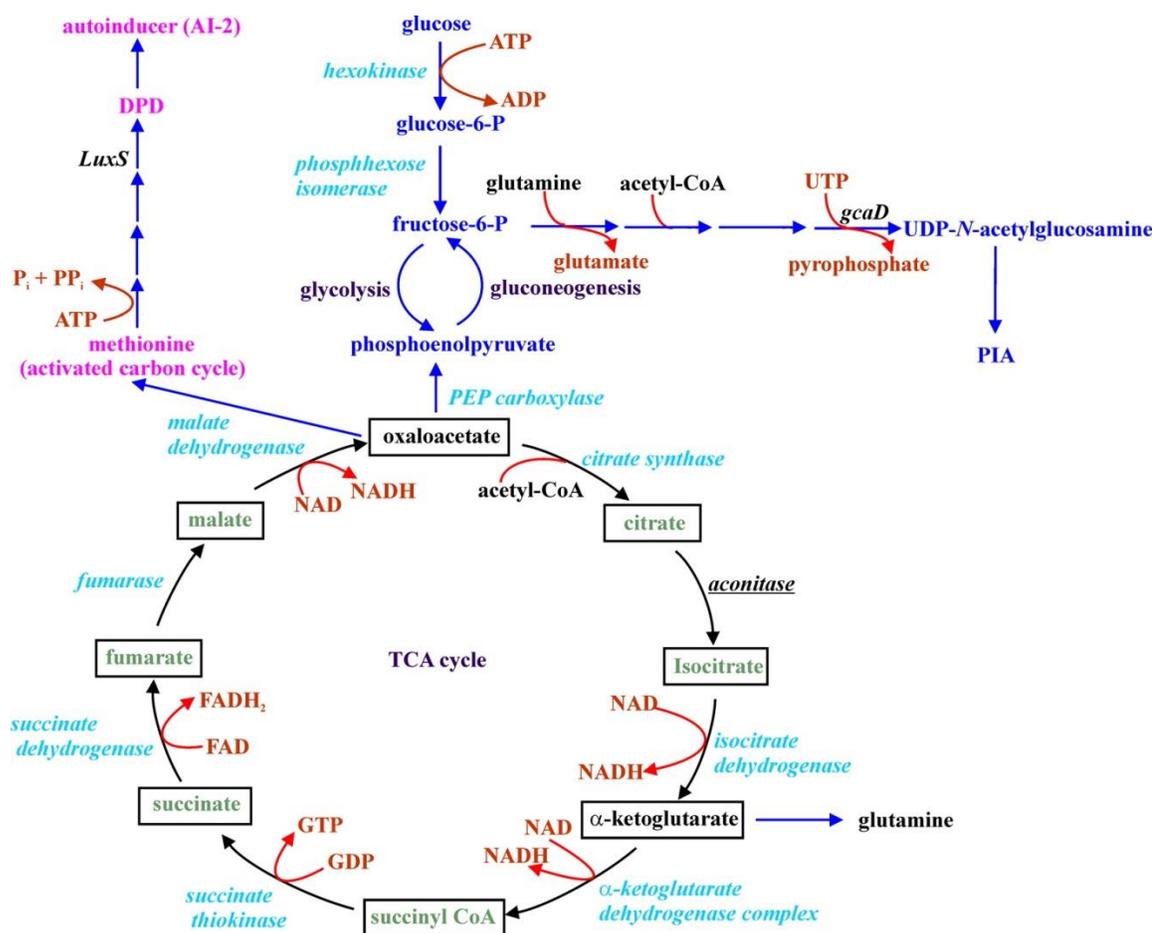


Figure 5.1 An illustration of the interrelationship of metabolic pathways associated with the TCA cycle and biofilm formation.

Our differential NMR metabolomics methodology has been applied to the study of *in vivo* drug activity in *Aspergillus nidulans*³⁵ and *Mycobacterium smegmatis*³⁶ 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 diagram³⁷ indicate that external biofilm signals inactivate the TCA cycle. Furthermore, a detailed analysis of the relative concentration changes of 55 different metabolites from 2D ¹H, ¹H TOCSY and 2D ¹H, ¹³C HSQC spectra implies the TCA cycle plays a central role in a metabolic signaling network.^{35,36,38,39} A metabolic network created with Cytoscape⁴⁰ 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.

5.2 Methods and Materials

5.2.1 Bacterial growth and NMR sample preparation

Staphylococcus epidermidis wild-type strain 1457 and the isogenic aconitase mutant strain 1457-*acnA::tetM* were grown in tryptic soy broth (TSB; BD Biosciences) without dextrose and supplemented with 0.25% glucose (Sigma Chemical) or 0.25% ¹³C₆-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°C with 225 rpm aeration in TSB or TSB supplemented with a stressor known 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) ¹H NMR experiments. 3 replicate bacterial cultures were grown for each bacterial strain or environmental condition for both the two-dimensional (2D) ¹H, ¹³C HSQC and the 2D ¹H, ¹H 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% ethanol²⁴ 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 ¹H NMR experiments, 2.74 O.D.₆₀₀ units were harvested for analysis and for both the 2D ¹H, ¹³C HSQC and 2D ¹H, ¹H 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% D₂O (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.

5.2.2 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 ^1H NMR spectra were collected using excitation sculpting⁴³ 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 ^1H 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 ^1H , ^{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 ^1H and ^{13}C dimensions, respectively. A total of 8 dummy scans and 128 scans were used to obtain each of the 2D ^1H , ^{13}C HSQC NMR spectra. 2D ^1H , ^1H 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 ^1H dimensions, respectively. A total of 16 dummy scans and 8 acquisition scans were used to obtain each of the 2D ^1H , ^1H TOCSY NMR spectra.

5.2.3 NMR Data Analysis

1D ^1H 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:

$$Z = \frac{X_i - \bar{X}}{\sigma} \quad (1)$$

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.⁴⁸ 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 NMRView⁴⁹ 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 three assigned peaks (*A*, *B*, *C*) in a 2D ^1H , ^{13}C HSQC spectrum. The 2D ^1H , ^{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 ^1H , ^{13}C HSQC and 2D ^1H , ^1H TOCSY spectra were assigned to specific metabolites using ^1H and ^{13}C chemical shift tolerances of 0.05 ppm and 0.50 ppm, respectively. Metabominer,⁵⁰ Madison Metabolomics Consortium Database (MMCD)⁵¹, the BioMagResBank,⁵² and Human Metabolome

Database⁵³ were used to identify metabolites. The presence of metabolites and metabolic pathways were verified with the KEGG⁵⁴ and Metacyc⁵⁵ 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.

5.3 Results and Discussion

5.3.1 NMR metabolomics and principal 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,⁵⁶ so variability in sample concentration was minimized, (iv) noise regions⁴⁸ 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.

5.3.2 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.^{16,57} 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 5.1.2a). 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.

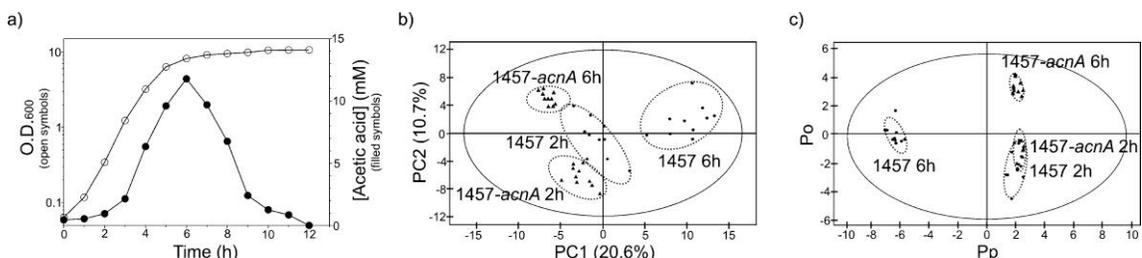


Figure 5.2 a) Typical *S. epidermidis* growth curve superimposed on the cellular production of acetic acid. b) 2D PCA scores plot and c) 2D OPLS-DA scores plot comparing 2 h growth of wild-type *S. epidermidis* 1457 (grey ●), 2 h growth of aconitase mutant strain 1457-*acnA::tetM* (grey ▲), 6 h growth of wild-type *S. epidermidis* 1457 (black ●), and 6 h growth of aconitase mutant strain 1457-*acnA::tetM* (black ▲). 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 (black ●) 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.

The growth phase-dependent activity of the TCA cycle in *S. epidermidis* is also apparent from the PCA 2D scores plot generated from 1D ^1H NMR spectra of *S. epidermidis* cell lysate (Figure 5.2b). 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 5.3) did not improve cluster separations. Each 1D ^1H 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 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.

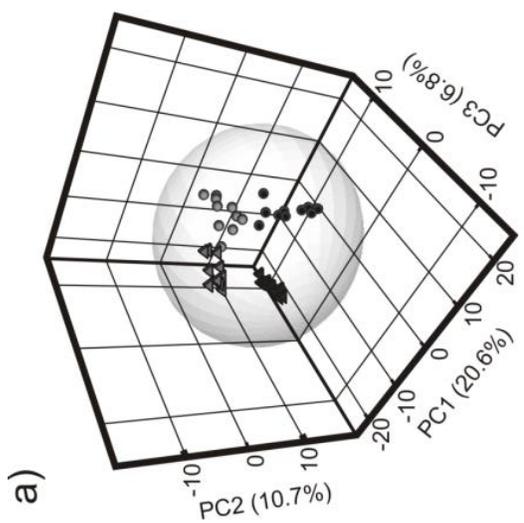
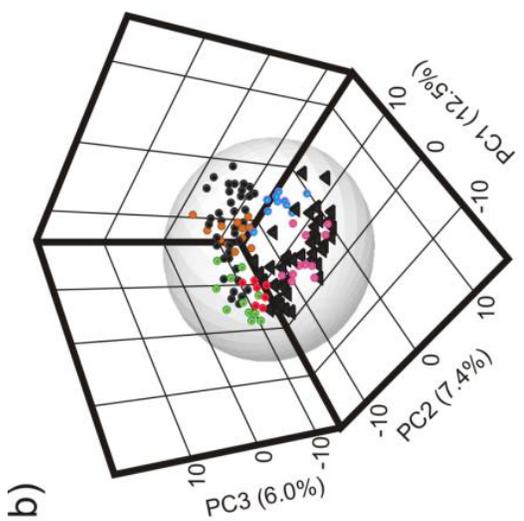
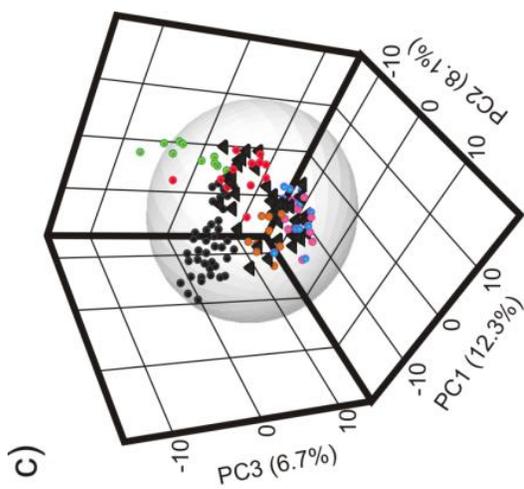
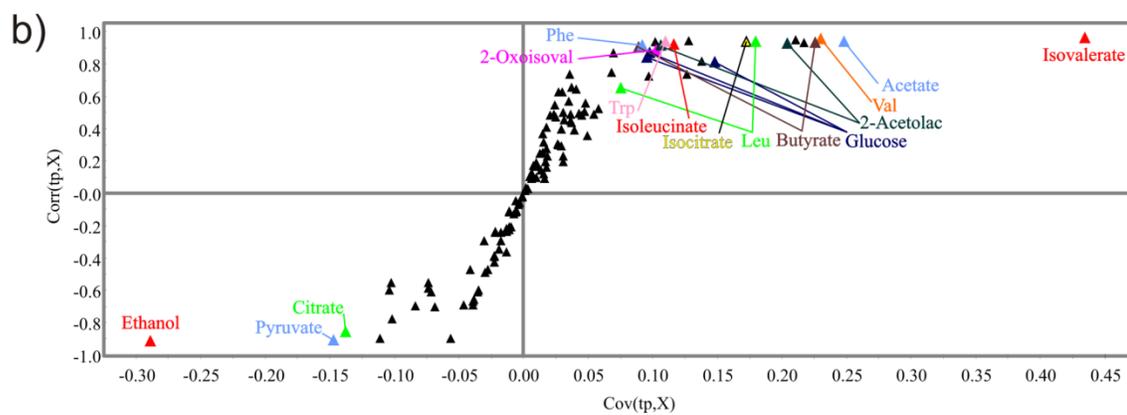
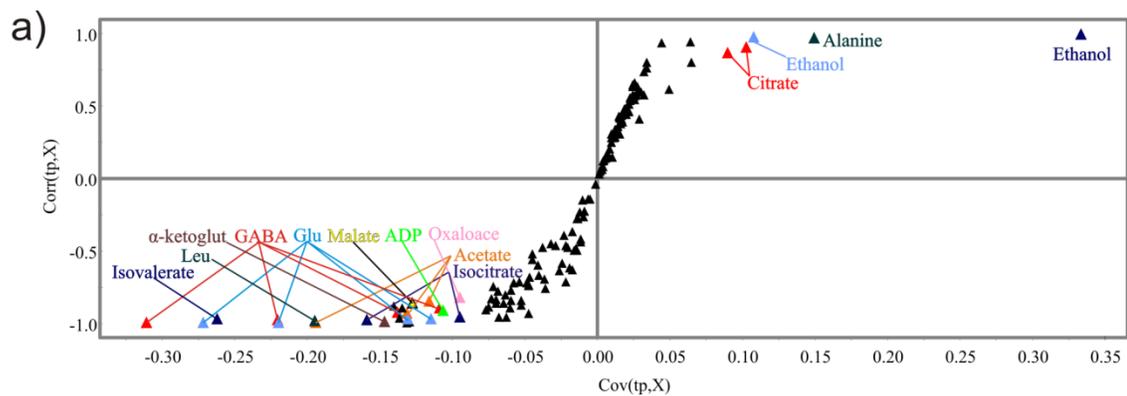


Figure 5.3. Three dimensional versions of the 2D PCA scores presented in Figure 5.2b, Figure 5.5b, and Figure 5.7b, respectively. **a)** 3D PCA scores plot comparing *S. epidermidis* 1457 strain (circles) and aconitase mutant (triangles) grown for 2 hours (grey) and 6 hours (black). The PCA model consists 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 $\mu\text{g}/\text{mL}$ tetracycline (●), with the addition of 5% NaCl (●), and 6 h growth of aconitase mutant strain 1457-*acnA::tetM* 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::tetM* in standard TSB media (▲), aconitase mutant strain 1457-*acnA::tetM* in iron-depleted media (DTSB) (●), with the addition of 4% ethanol (●), with the addition of 2% glucose (●), with the addition of 0.06 $\mu\text{g}/\text{mL}$ 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.

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 5.2c). 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 5.4) generated from the OPLS-DA provides unambiguous identification of the major contributors to the class separation (i.e., 1H

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.



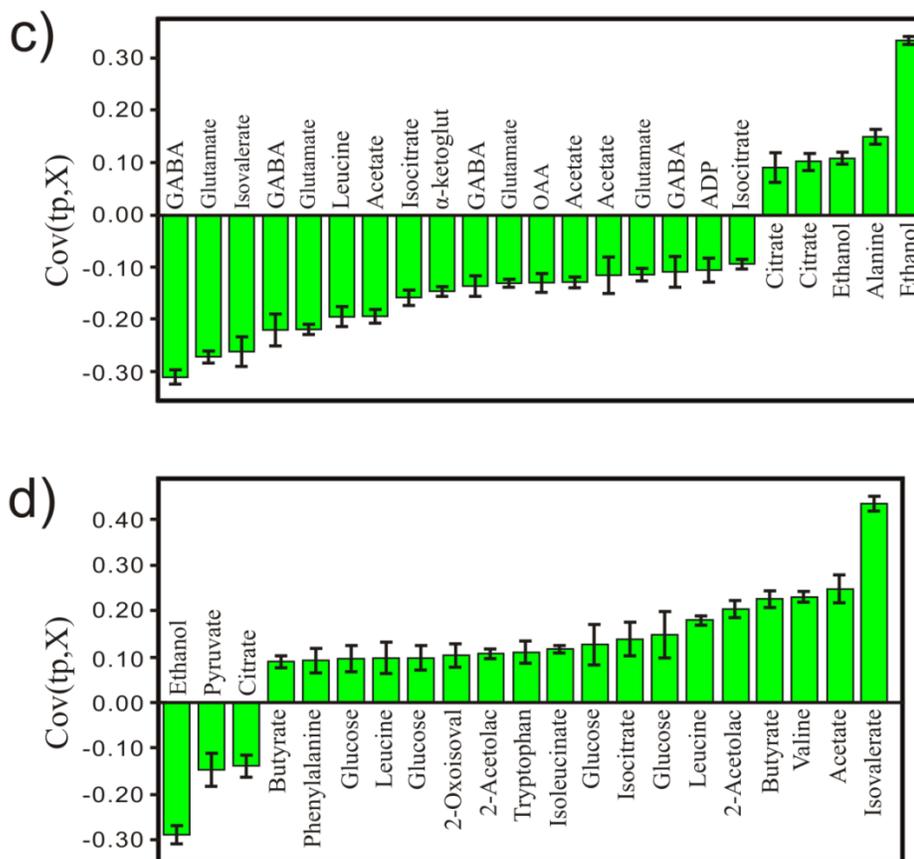


Figure 5.4 a) OPLS-DA S-plots comparing the *S. epidermidis* 1457 and aconitase mutant strain 1457-*acnA::tetM* where both cell cultures were grown for 6 hours. The two cell cultures were shown to be separated along the PC1 axis in Figure 5.2a. Each point in the S-plot represents a specific bin containing 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::tetM* grown for 2 hours and 6 hours. The two conditions are shown to be separated along the PC2 axis in Figure 5.2a. 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::tetM* 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 5.10 from the analysis of 2D NMR data. **d)** OPLS-DA loading plot comparing the mutant strain 1457-*acnA::tetM* grown for 2 hours and 6 hours.

5.3.3 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.⁵⁸ This process requires that *S. epidermidis* “sense” changes in its environment and the availability of nutrients, such as changes in temperature, O₂ levels, osmolarity, ethanol, glucose and iron.^{29,31,33,59-63} It is reasonable to expect that different external factors would trigger distinct signaling 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 5.1). 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

28,33,64

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,⁶⁵ which has been shown to be a regulator of *icaR*,⁶⁶ a negative regulator of the *icaADBC* operon that is required for PIA synthesis and biofilm formation.⁶⁷ Subinhibitory concentrations

of antibiotics enhance *icaADBC* gene expression^{28,68} by potentially inhibiting TcaR, a weak negative regulator of *icaADBC* gene expression.⁶⁹ AI-2 is an intercellular signaling molecule that has a modest effect on staphylococcal biofilms.⁶⁴ 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 5.5a) and the associated metabolic tree (Figure 5.5c) 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 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 5.12). 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.^{70,71}

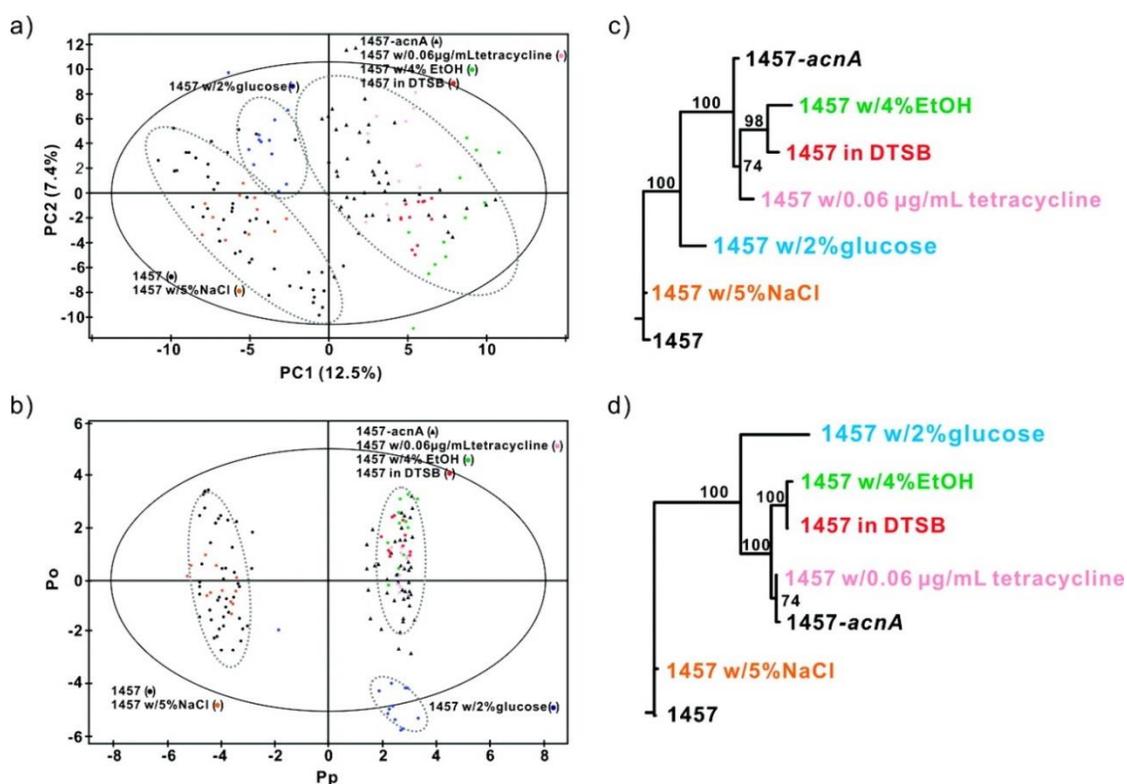


Figure 5.5 (a) 2D PCA scores plot and (b) 2D OPLS-DA comparing wild-type *S. epidermidis* 1457 cells grown 6 h in standard TSB media (black ●), with *S. epidermidis* 1457 cells grown 6 h in iron-depleted media (DTSB) (red ●), with the addition of 4% ethanol (green ●), with the addition of 2% glucose (blue ●), with the addition of 0.06 μg/mL tetracycline (pink ●), with the addition of 5% NaCl (orange ●), and 6 h growth of aconitase mutant strain 1457-*acnA*::*tetM* in standard TSB media (black ▲). 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 (black ●) 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 R2X of 0.637, R2Y of 0.966 and Q2 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.

The PCA results were used to guide a subsequent analysis using OPLS-DA (Figure 5.5b) and the corresponding metabolomics tree diagram (Figure 5.5d). 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 $\mu\text{g/mL}$ tetracycline or iron-limitation again formed a single 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.

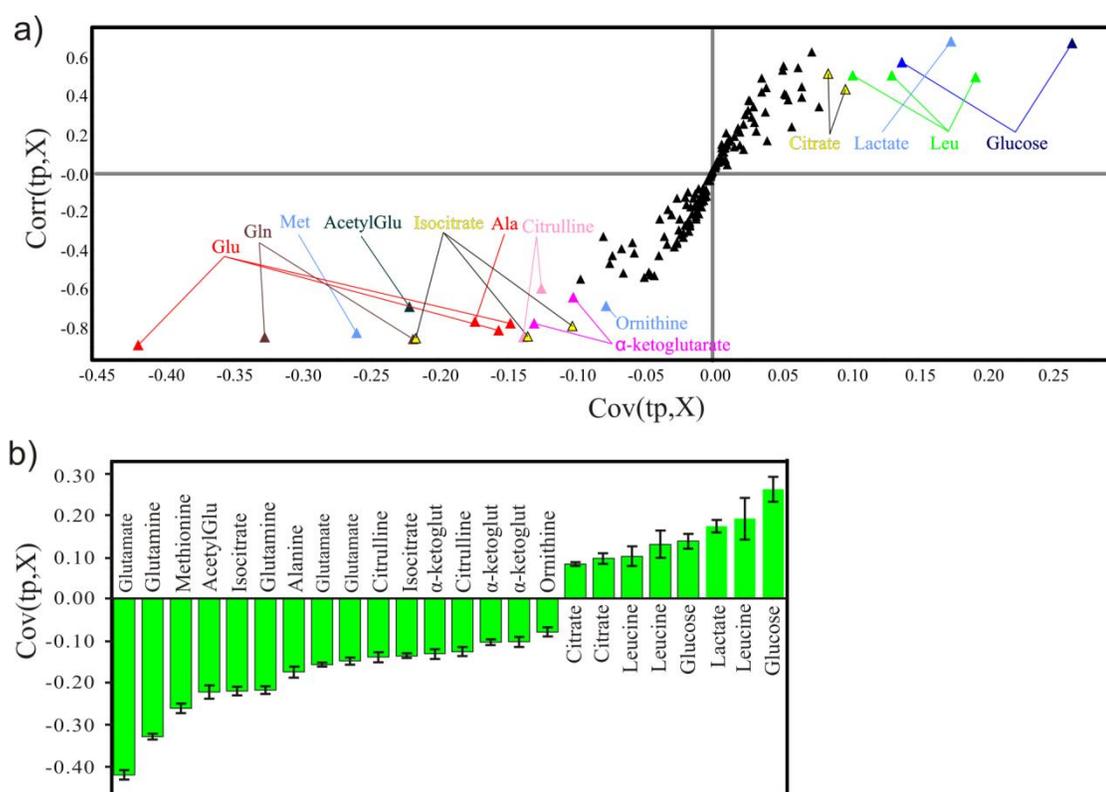


Figure 5.6 a) OPLS-DA S-plots comparing wild-type *S. epidermidis* 1457 and *S. epidermidis* cultured under the different stress conditions described in Figure 5.5. b) OPLS-DA loading plot showing the contribution of the identified metabolites from the S-plot. These results are comparable to the bar graphs depicted in Figure 5.10 from the analysis of 2D NMR data.

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 $\mu\text{g/mL}$ 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 5.7a) and metabolomic tree (Figure 5.7c) 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 5.7b) and the corresponding metabolomics tree diagram (Figure 5.7d). 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 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.

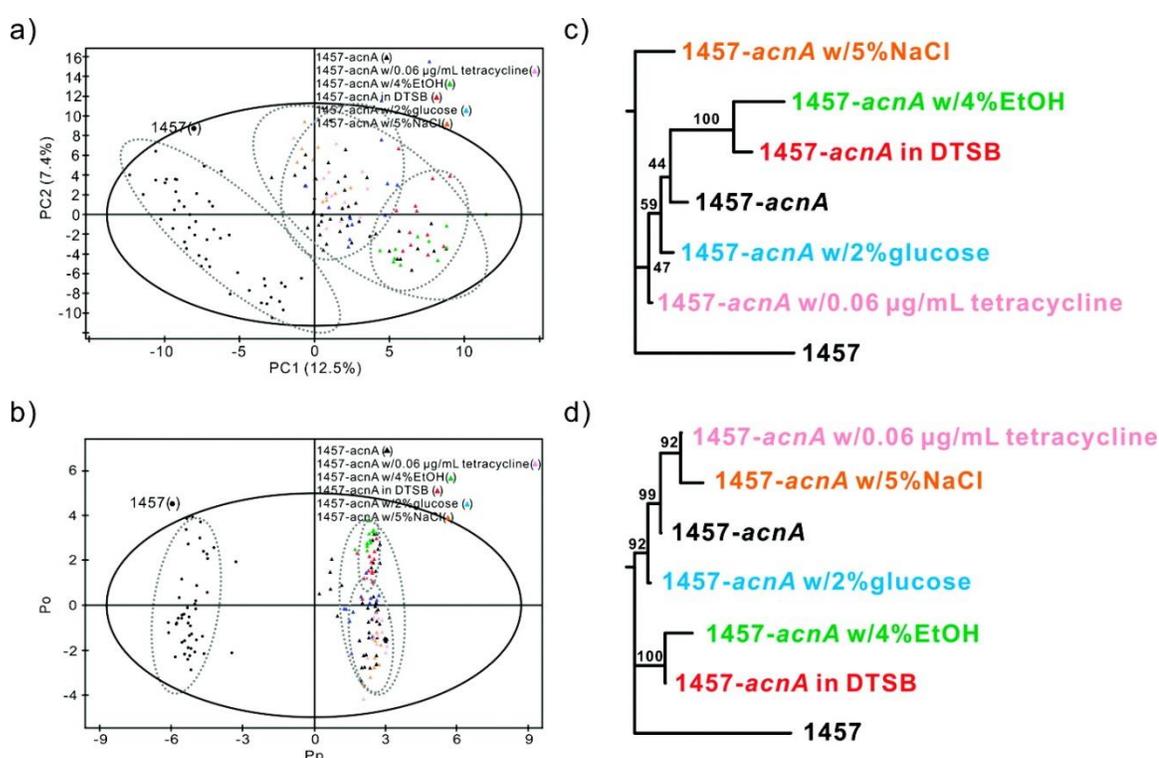


Figure 5.7 (a) 2D PCA scores plot and (b) 2D OPLS-DA comparing wild-type *S. epidermidis* 1457 cells grown 6 h in standard TSB media (black ●), 6 h growth of aconitase mutant strain 1457-*acnA*::*tetM* in standard TSB media (black ▲), aconitase mutant strain 1457-*acnA*::*tetM* in iron-depleted media (DTSB) (red ▲), with the addition of 4% ethanol (green ▲), with the addition of 2% glucose (blue ▲), with the addition of 0.06 µg/mL tetracycline (pink ▲), and with the addition of 5% NaCl (orange ▲). 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::tetM* in standard TSB media, (iii) aconitase mutant strain 1457-*acnA::tetM* in DTSB or with the addition of 4% ethanol, and (iv), aconitase mutant strain 1457-*acnA::tetM* with the addition of 2% glucose, 0.06 $\mu\text{g/mL}$ tetracycline, or 5% NaCl. For the OPLS-DA scores plot, the 6 h growth of wild-type *S. epidermidis* 1457 (black ●) 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.

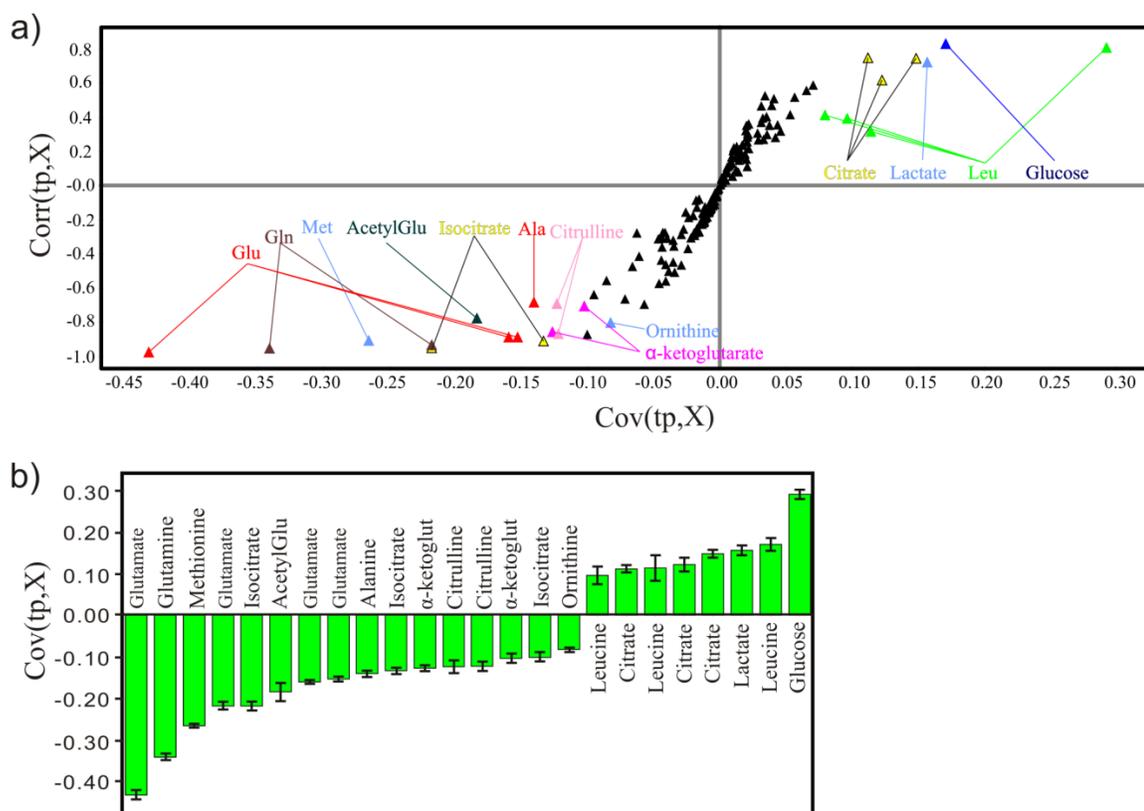


Figure 5.8 a) OPLS-DA S-plots comparing the *S. epidermidis* 1457 and the aconitase mutant 1457-*acnA::tetM* under the different stress conditions described in Figure 5.7b. b) OPLS-DA loading plot showing the contribution of the identified metabolites from the S-plot. These results are comparable to the bar graphs depicted in Figure 5.10 from the analysis of 2D NMR data.

5.3.4 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 through 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 ¹H, ¹³C HSQC and 2D ¹H, ¹H TOCSY NMR spectra was applied to quantitate metabolite changes in the *S. epidermidis* metabolome caused by Fe-limitation and 4% ethanol (Figure 5.9).²³

Figure 5.9 Overlay of (a) 2D ^1H , ^{13}C HSQC spectra and (b) 2D ^1H , ^1H TOCSY spectra comparing wild-type *S. epidermidis* strain 1457 (red) and aconitase mutant strain 1457-*acnA::tetM* (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.

2D NMR spectra improve metabolite identification by reducing the complexity and congestion of 1D ^1H NMR spectrum by spreading the information into two-dimensions. Additionally, the 2D ^1H , ^{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 ^1H , ^1H 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 ^1H , ^{13}C HSQC spectrum in addition to non-isotope labeled metabolites produced from other carbon sources. Therefore, the 2D ^1H , ^{13}C HSQC and 2D ^1H , ^1H 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 for both the 2D ^1H , ^{13}C HSQC and 2D ^1H , ^1H 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 ^1H , ^{13}C HSQC and 2D ^1H , ^1H 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 5.10. 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 3, 4). The OPLS-DA S-plots (Figure 5.8) identifies the relative contribution of each bin (^1H NMR chemical shift) to the clustering in the corresponding 2D scores plot. Each NMR bin with a high reliability ($p(\text{corr})[1] \sim 1$ 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 5.10 bar graph.

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 5.9. 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 5.2b). 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.⁷² Aside from the glycolytic pathway, the NMR data indicates that wild-type bacteria tend to utilize glucose more efficiently based on small decreases in amino sugar and aromatic metabolites.

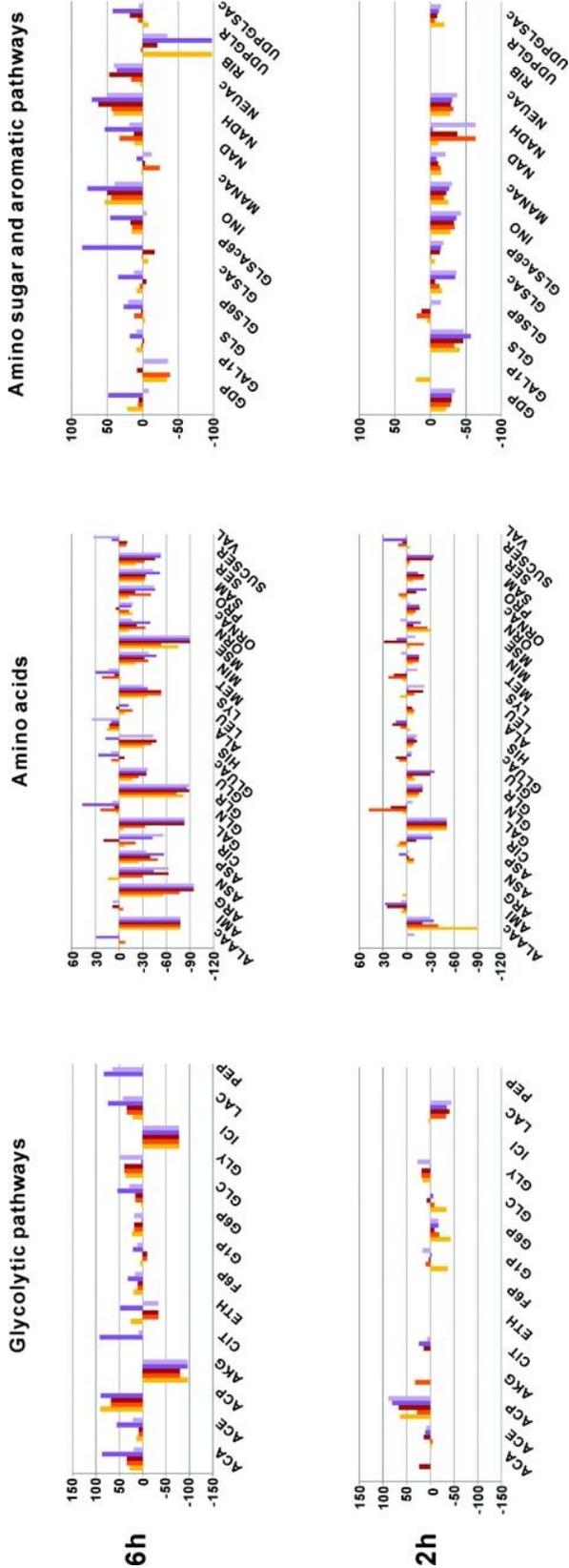


Figure 5.10 Bar graphs showing the percent change in metabolite concentrations relative to wild-type *S. epidermidis* strain 1457 grown in standard TSB media. Metabolite concentration changes were measured after 2 and 6 h bacterial growths for the aconitase mutant strain 1457-*acnA::tetM* in TSB media (light yellow), aconitase mutant strain 1457-*acnA::tetM* with iron-depletion (dark yellow), aconitase mutant strain 1457-*acnA::tetM* with the addition of 4% ethanol (orange), wild-type *S. epidermidis* strain 1457 with iron-depletion (dark purple), and wild-type *S. epidermidis* strain 1457 with the addition of 4% ethanol (light purple). 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-glucuronate), UDPGLSAc (UDP-NAc-d-glucosamine), VAL (Valine).

5.3.5 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 5.11) was generated by manually associating each metabolite to its corresponding pathway from the KEGG⁵⁴ database and then using the automated biological network modules integrated into Cytoscape.^{73,74} 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.

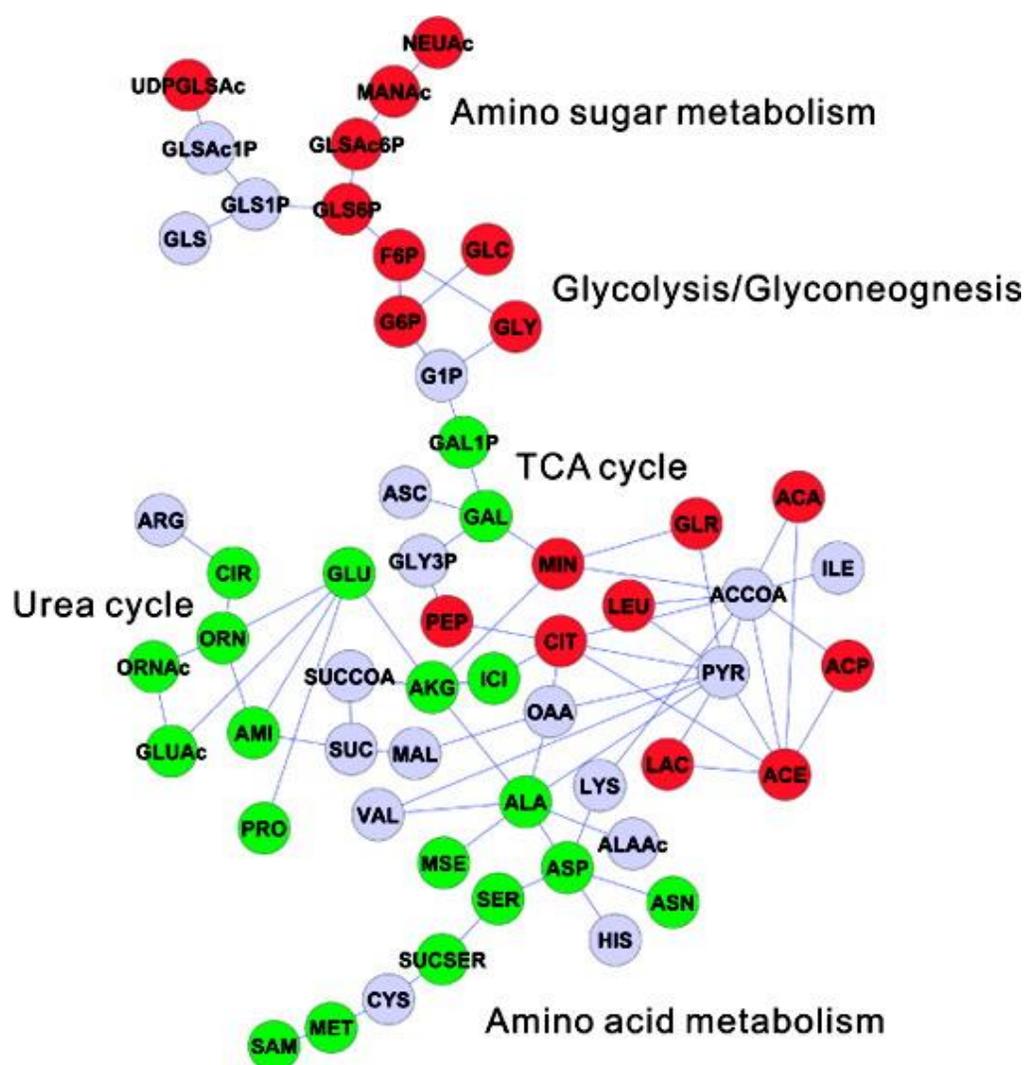


Figure 5.11 Cytoscape network depicting the metabolite concentration changes caused by the inactivation of the TCA cycle. Nodes colored red correspond to metabolites with an increase in concentration due to TCA inactivation. Nodes colored green correspond to metabolites with a decrease in concentration due to TCA inactivation. Nodes colored gray 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 labelled on the network. The metabolite names were abbreviated as described in the legend to Figure 5.10.

5.4 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.

5.5 References

- (1) Brent, R. *FEBS Lett.* **2009**, 583, 4019.

- (2) Jorgensen, C.; Linding, R. *Curr. Opin. Genet. Dev.*, **20**, 15.
- (3) Kholodenko, B. N. *FEBS Lett.* **2009**, *583*, 4006.
- (4) Scott, J. D.; Pawson, T. *Science* **2009**, *326*, 1220.
- (5) Mascher, T.; Helmann, J. D.; Uden, G. *Microbiol. Mol. Biol. Rev.* **2006**, *70*, 910.
- (6) Gill, S. R.; Fouts, D. E.; Archer, G. L.; Mongodin, E. F.; DeBoy, R. T.; Ravel, J.; Paulsen, I. T.; Kolonay, J. F.; Brinkac, L.; Beanan, M.; Dodson, R. J.; Daugherty, S. C.; Madupu, R.; Angiuoli, S. V.; Durkin, A. S.; Haft, D. H.; Vamathevan, J.; Khouri, H.; Utterback, T.; Lee, C.; Dimitrov, G.; Jiang, L.; Qin, H.; Weidman, J.; Tran, K.; Kang, K.; Hance, I. R.; Nelson, K. E.; Fraser, C. M. *J. Bacteriol.* **2005**, *187*, 2426.
- (7) Novick, R. P.; Geisinger, E. *Annu. Rev. Genet.* **2008**, *42*, 541.
- (8) Reznikoff, W. S.; Siegele, D. A.; Cowing, D. W.; Gross, C. A. *Annu. Rev. Genet.* **1985**, *19*, 355.
- (9) Wu, S.; de, L. H.; Tomasz, A. *J. Bacteriol.* **1996**, *178*, 6036.
- (10) Senn, M. M.; Giachino, P.; Homerova, D.; Steinhuber, A.; Strassner, J.; Kormanec, J.; Fluckiger, U.; Berger-Bachi, B.; Bischoff, M. *J. Bacteriol.* **2005**, *187*, 8006.
- (11) Somerville, G. A.; Proctor, R. A. *Microbiol. Mol. Biol. Rev.* **2009**, *73*, 233.
- (12) Lloyd, D. *Process Biochem.* **1966**, *1*, 465.
- (13) Bae, T.; Banger, A. K.; Wallace, A.; Glass, E. M.; Aslund, F.; Schneewind, O.; Missiakas, D. M. *Proc Natl Acad Sci U S A* **2004**, *101*, 12312.

- (14) Begun, J.; Sifri, C. D.; Goldman, S.; Calderwood, S. B.; Ausubel, F. M. *Infect Immun* **2005**, *73*, 872.
- (15) Coulter, S. N.; Schwan, W. R.; Ng, E. Y.; Langhorne, M. H.; Ritchie, H. D.; Westbrook-Wadman, S.; Hufnagle, W. O.; Folger, K. R.; Bayer, A. S.; Stover, C. K. *Mol Microbiol* **1998**, *30*, 393.
- (16) Sadykov, M. R.; Olson, M. E.; Halouska, S.; Zhu, Y.; Fey, P. D.; Powers, R.; Somerville, G. A. *J Bacteriol* **2008**, *190*, 7621.
- (17) Cramton, S. E.; Gerke, C.; Schnell, N. F.; Nichols, W. W.; Gotz, F. *Infect Immun* **1999**, *67*, 5427.
- (18) Maira-Litran, T.; Kropec, A.; Abeygunawardana, C.; Joyce, J.; Mark, G., 3rd; Goldmann, D. A.; Pier, G. B. *Infect Immun* **2002**, *70*, 4433.
- (19) Rupp, M. E.; Ulphani, J. S.; Fey, P. D.; Bartscht, K.; Mack, D. *Infect Immun* **1999**, *67*, 2627.
- (20) Rupp, M. E.; Ulphani, J. S.; Fey, P. D.; Mack, D. *Infect Immun* **1999**, *67*, 2656.
- (21) Fitzpatrick, F.; Humphreys, H.; O'Gara, J. P. *Clin Microbiol Infect* **2005**, *11*, 967.
- (22) Vuong, C.; Kidder, J. B.; Jacobson, E. R.; Otto, M.; Proctor, R. A.; Somerville, G. A. *J Bacteriol* **2005**, *187*, 2967.
- (23) Sadykov, M. R.; Zhang, B.; Halouska, S.; Nelson, J. L.; Kreimer, L. W.; Zhu, Y.; Powers, R.; Somerville, G. A. *J. Biol. Chem.* **2010**, *285*, 36616.
- (24) Knobloch, J. K. M.; Bartscht, K.; Sabottke, A.; Rohde, H.; Feucht, H.-H.; Mack, D. *Journal of Bacteriology* **2001**, *183*, 2624.

- (25) Campbell, I. M.; Crozier, D. N.; Pawagi, A. B.; Buivids, I. A. *Journal of Clinical Microbiology* **1983**, *18*, 408.
- (26) Mack, D.; Siemssen, N.; Laufs, R. *Infection and Immunity* **1992**, *60*, 2048.
- (27) Gerke, C.; Kraft, A.; Sussmuth, R.; Schweitzer, O.; Gotz, F. *Journal of Biological Chemistry* **1998**, *273*, 18586.
- (28) Rachid, S.; Ohlsen, K.; Witte, W.; Hacker, J.; Ziebuhr, W. *Antimicrob Agents Chemother* **2000**, *44*, 3357.
- (29) Cramton, S. E.; Ulrich, M.; Gotz, F.; Doring, G. *Infect Immun* **2001**, *69*, 4079.
- (30) Deighton, M.; Borland, R. *Infection and Immunity* **1993**, *61*, 4473.
- (31) Elci, S.; Atmaca, S.; Gul, K. *Cytobios* **1995**, *84*, 141.
- (32) Evans, E.; Brown, M. R. W.; Gilbert, P. *Microbiology (Reading, United Kingdom)* **1994**, *140*, 153.
- (33) Rachid, S.; Ohlsen, K.; Wallner, U.; Hacker, J.; Hecker, M.; Ziebuhr, W. *J Bacteriol* **2000**, *182*, 6824.
- (34) Zhao, L.; Xue, T.; Shang, F.; Sun, H.; Sun, B. *Infect. Immun.* **2010**, *78*, 3506.
- (35) Forgue, P.; Halouska, S.; Werth, M.; Xu, K.; Harris, S.; Powers, R. *Journal of Proteome Research* **2006**, *5*, 1916.
- (36) Halouska, S.; Chacon, O.; Fenton, R. J.; Zinniel, D. K.; Barletta, R. G.; Powers, R. *Journal of Proteome Research* **2007**, *6*, 4608.

- (37) Werth, M. T.; Halouska, S.; Shortridge, M. D.; Zhang, B.; Powers, R. *Analytical Biochemistry* **2010**, 399, 56.
- (38) Lindon, J. C.; Holmes, E.; Nicholson, J. K. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2001**, 39, 40.
- (39) Stoyanova, R.; Brown, T. R. *NMR Biomed* **2001**, 14, 271.
- (40) Killcoyne, S.; Carter, G. W.; Smith, J.; Boyle, J. *Methods Mol Biol* **2009**, 563, 219.
- (41) Cao, J. G.; Meighen, E. A. *J. Biol. Chem.* **1989**, 264, 21670.
- (42) Semmelhack, M. F.; Campagna, S. R.; Federle, M. J.; Bassler, B. L. *Org. Lett.* **2005**, 7, 569.
- (43) Nguyen, B. D.; Meng, X.; Donovan, K. J.; Shaka, A. J. *J Magn Reson* **2007**, 184, 263.
- (44) Palmer, A. G. I.; Cavanagh, J.; Wright, P. E.; Rance, M. J. *Journal of Magnetic Resonance* **1991**, 93, 20.
- (45) Kay, L. E.; Keifer, P.; Saarinen, T. *J Am Chem Soc* **1992**, 114, 3.
- (46) Bax, A.; Davis, D. G. *Journal of Magnetic Resonance* **1985**, 65, 6.
- (47) Piotto, M.; Saudek, V.; Sklenar, V. *J Biomol NMR* **1992**, 2, 661.
- (48) Halouska, S.; Powers, R. *Journal of Magnetic Resonance* **2006**, 178, 88.
- (49) Johnson, B. A. *Methods Mol Biol* **2004**, 278, 313.

- (50) Xia, J.; Bjorndahl, T. C.; Tang, P.; Wishart, D. S. *BMC Bioinf.* **2008**, *9*, No pp. given.
- (51) Cui, Q.; Lewis, I. A.; Hegeman, A. D.; Anderson, M. E.; Li, J.; Schulte, C. F.; Westler, W. M.; Eghbalian, H. R.; Sussman, M. R.; Markley, J. L. *Nat Biotechnol* **2008**, *26*, 162.
- (52) Ulrich, E. L.; Akutsu, H.; Doreleijers, J. F.; Harano, Y.; Ioannidis, Y. E.; Lin, J.; Livny, M.; Mading, S.; Maziuk, D.; Miller, Z.; Nakatani, E.; Schulte, C. F.; Tolmie, D. E.; Kent, W. R.; Yao, H.; Markley, J. L. *Nucleic Acids Res.* **2008**, *36*, D402.
- (53) Wishart, D. S.; Knox, C.; Guo, A. C.; Eisner, R.; Young, N.; Gautam, B.; Hau, D. D.; Psychogios, N.; Dong, E.; Bouatra, S.; Mandal, R.; Sinelnikov, I.; Xia, J.; Jia, L.; Cruz, J. A.; Lim, E.; Sobsey, C. A.; Shrivastava, S.; Huang, P.; Liu, P.; Fang, L.; Peng, J.; Fradette, R.; Cheng, D.; Tzur, D.; Clements, M.; Lewis, A.; De, S. A.; Zuniga, A.; Dawe, M.; Xiong, Y.; Clive, D.; Greiner, R.; Nazyrova, A.; Shaykhutdinov, R.; Li, L.; Vogel, H. J.; Forsythe, I. *Nucleic Acids Res.* **2009**, *37*, D603.
- (54) Kanehisa, M.; Araki, M.; Goto, S.; Hattori, M.; Hirakawa, M.; Itoh, M.; Katayama, T.; Kawashima, S.; Okuda, S.; Tokimatsu, T.; Yamanishi, Y. *Nucleic Acids Res.* **2008**, *36*, D480.
- (55) Caspi, R.; Altman, T.; Dale, J. M.; Dreher, K.; Fulcher, C. A.; Gilham, F.; Kaipa, P.; Karthikeyan, A. S.; Kothari, A.; Krummenacker, M.; Latendresse, M.; Mueller, L. A.; Paley, S.; Popescu, L.; Pujar, A.; Shearer, A. G.; Zhang, P.; Karp, P. D. *Nucleic Acids Res.* **2010**, *38*, D473.
- (56) van den Berg, R. A.; Hoefsloot, H. C.; Westerhuis, J. A.; Smilde, A. K.; van der Werf, M. J. *BMC Genomics* **2006**, *7*, 142.
- (57) Somerville, G. A.; Chaussee, M. S.; Morgan, C. I.; Fitzgerald, J. R.; Dorward, D. W.; Reitzer, L. J.; Musser, J. M. *Infect. Immun.* **2002**, *70*, 6373.
- (58) Mack, D.; Becker, P.; Chatterjee, I.; Dobinsky, S.; Knobloch, J. K. M.; Peters, G.; Rohde, H.; Herrmann, M. *Int. J. Med. Microbiol.* **2004**, *294*, 203.

- (59) Collins, F. M.; Lascelles, J. *J Gen Microbiol* **1962**, *29*, 531.
- (60) Costerton, J. W.; Stewart, P. S.; Greenberg, E. P. *Science* **1999**, *284*, 1318.
- (61) Kornmann, H.; Duboc, P.; Niederberger, P.; Marison, I.; von Stockar, U. *Appl Microbiol Biotechnol* **2003**, *62*, 168.
- (62) Throup, J. P.; Zappacosta, F.; Lunsford, R. D.; Annan, R. S.; Carr, S. A.; Lonsdale, J. T.; Bryant, A. P.; McDevitt, D.; Rosenberg, M.; Burnham, M. K. *Biochemistry* **2001**, *40*, 10392.
- (63) Tomlins, R. I.; Pierson, M. D.; Ordal, Z. J. *Can J Microbiol* **1971**, *17*, 759.
- (64) Miller, M. B.; Bassler, B. L. *Annu. Rev. Microbiol.* **2001**, *55*, 165.
- (65) Lim, Y.; Jana, M.; Luong, T. T.; Lee, C. Y. *J Bacteriol* **2004**, *186*, 722.
- (66) Cue, D.; Lei, M. G.; Luong, T. T.; Kuechenmeister, L.; Dunman, P. M.; O'Donnell, S.; Rowe, S.; O'Gara, J. P.; Lee, C. Y. *J. Bacteriol.* **2009**, *191*, 6363.
- (67) Heilmann, C.; Schweitzer, O.; Gerke, C.; Vanittanakom, N.; Mack, D.; Goetz, F. *Mol. Microbiol.* **1996**, *20*, 1083.
- (68) Hoffman, L. R.; D'Argenio, D. A.; MacCoss, M. J.; Zhang, Z.; Jones, R. A.; Miller, S. I. *Nature (London, U. K.)* **2005**, *436*, 1171.
- (69) Chang, Y.-M.; Jeng, W.-Y.; Ko, T.-P.; Yeh, Y.-J.; Chen, C. K. M.; Wang, A. H. J. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 8617.
- (70) Sousa, C.; Henriques, M.; Azeredo, J.; Teixeira, P.; Oliveira, R. *World J. Microbiol. Biotechnol.* **2008**, *24*, 423.
- (71) Croes, S.; Deurenberg, R. H.; Boumans, M.-L. L.; Beisser, P. S.; Neef, C.; Stobberingh, E. E. *BMC Microbiol.* **2009**, *9*, No pp. given.

(72) Somerville, G. A.; Said-Salim, B.; Wickman, J. M.; Raffel, S. J.; Kreiswirth, B. N.; Musser, J. M. *Infect Immun* **2003**, *71*, 4724.

(73) Cline, M. S.; Smoot, M.; Cerami, E.; Kuchinsky, A.; Landys, N.; Workman, C.; Christmas, R.; Avila-Campilo, I.; Creech, M.; Gross, B.; Hanspers, K.; Isserlin, R.; Kelley, R.; Killcoyne, S.; Lotia, S.; Maere, S.; Morris, J.; Ono, K.; Pavlovic, V.; Pico, A. R.; Vailaya, A.; Wang, P. L.; Adler, A.; Conklin, B. R.; Hood, L.; Kuiper, M.; Sander, C.; Schmulevich, I.; Schwikowski, B.; Warner, G. J.; Ideker, T.; Bader, G. D. *Nat Protoc* **2007**, *2*, 2366.

(74) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. *Genome Res* **2003**, *13*, 2498.

(75) Li, M.; Villaruz, A. E.; Vadyvaloo, V.; Sturdevant, D. E.; Otto, M. *BMC Microbiol* **2008**, *8*, 4.

(76) Feder, M. E.; Walser, J. C. *J. Evol. Biol.* **2005**, *18*, 901.

APPENDIX A: EFFECT OF AUTOINDUCER-2 (AI-2) ON THE *S. EPIDERMIDIS* METABOLOME

Autoinducer 2 (AI-2) is believed to be an intercellular signaling molecule used in quorum sensing by both gram-negative and gram-positive bacteria.⁶⁴ AI-2 minimally represses biofilm formation in *S. epidermidis* where it has been reported that AI-2 indirectly effects the transcription of metabolic genes.⁷⁵ This raises the possibility that AI-2 may also function by affecting TCA cycle activity. To investigate the impact of AI-2 on the TCA cycle, bacterial cultures of wild-type *S. epidermidis* and the corresponding aconitase mutant strain were treated with AI-2. Zhao *et al.*³⁴ recently demonstrated that *S. aureus* cultures supplemented with 390 nM of 4,5-dihydroxy-2,3-pentanedione (DPD), the AI-2 precursor molecule, restored a *luxS* deletion mutant. Thus, 400 nM of AI-2 was used for our NMR metabolomics studies since this concentration of AI-2 was identified as physiologically relevant. The 2D PCA and 2D OPLS-DA scores plots comparing AI-2 treated and untreated bacterial cultures indicates AI-2 did not change the typical clustering pattern (Figure 5.12). In other words, the NMR metabolomics data suggests AI-2 does not affect TCA activity or cause a corresponding perturbation in the *S. epidermidis* metabolome. This observation is in contrast to the data reported by Li *et al.*, where AI-2 was shown to effect the transcription of numerous metabolic genes.⁷⁵ The apparent discrepancy highlights the benefits of NMR metabolomics, where changes in the metabolome reflect actual changes in protein activity. Conversely, a change in a transcriptional profile does not necessarily correlate with a difference in protein activity.⁷⁶

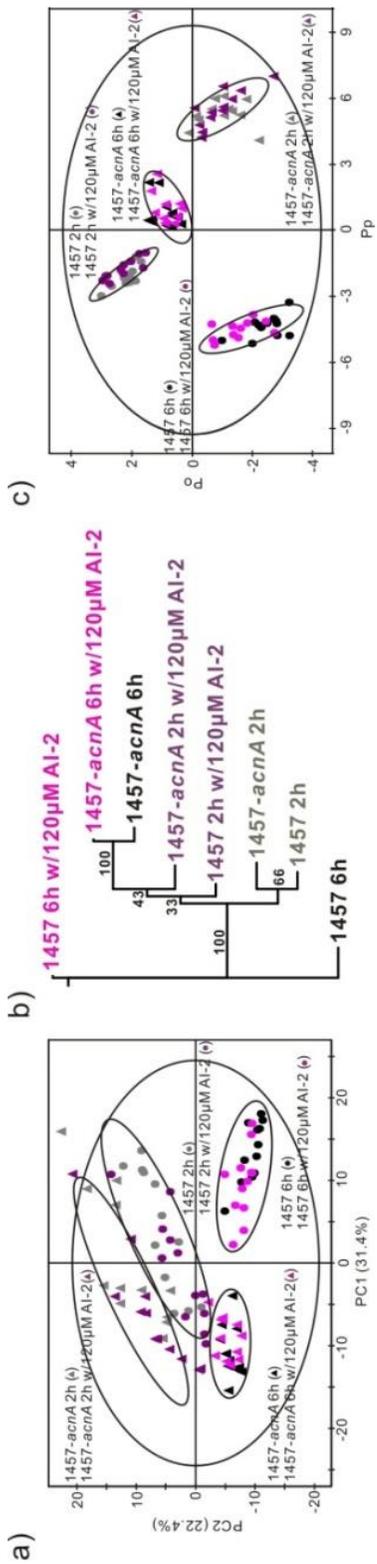


Figure 5.12 2D PCA scores plot comparing wild-type *S. epidermidis* strain 1457 and aconitase mutant strain 1457-*acnA::tetM* with and without the addition of 400 nM of AI-2. Cells were grown for 2 h or 6 h in standard TSB media, 2 h growth of wild-type *S. epidermidis* strain 1457 with AI-2 (●) and without AI-2 (●); aconitase mutant strain 1457-*acnA::tetM* 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::tetM* 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::tetM* 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

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 sigma factors, mechanosensors, small RNAs, riboswitches, and 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.) provide biosynthetic intermediates, ii.) generate reducing potential, and iii.) 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 Methods and Materials

6.2.1 Bacterial strains, media, and growth conditions

S. epidermidis wild-type strain 1457¹⁸ 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) technique²⁰ to replace the gene of interest with an antibiotic resistance marker (*i.e.*, *ermB* or *tetM*). Primers were designed to amplify approximately 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; BD Biosciences) supplemented with 0.25% glucose (Sigma Chemical) or 0.25% ¹³C₆-glucose (Cambridge Isotope Laboratories). Defferated TSB (DTSB) was prepared by adding 50 g of Chelex 100 (Sigma Chemical) to approximately 1 L of TSB and stirring at 4 °C for 20 h. After 20 h, the Chelex resin was removed, 1 mM MgSO₄ was added, the volume was adjusted to 1 L, 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 (OD₆₀₀). Antibiotics, when used, were purchased from Fisher Scientific or Sigma Chemical and used at the following concentrations: chloramphenicol (8 µg/mL), trimethoprim (10 g/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 M fluorocitrate. The samples were lysed in 2 mL screw cap tubes containing lysing matrix B using a FastPrep instrument (MP Biomedical). 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 $A_{240} \text{ min}^{-1}$ 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 primers:

femC- forward, 5'-GATGTTTGATGGTTCATCTATTGAAGGTTTCG-3';

femC-reverse, 5'-GCAGTATCAGTCAATTGTAAATCACCTTCAG-3';

RNAIII-forward, 5'-TGAAAAATTTGCTTAATCTAGTCGAGTG-3';

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 (1D) ^1H spectra were prepared from 10 independent, 25 mL *S. epidermidis* cultures. Two-dimensional (2D) ^1H , ^{13}C HSQC^{25,26} and 2D ^1H , ^1H TOCSY²⁷ spectra were prepared from 3 independent 50 mL cultures. The TSB medium used in the 2D ^1H , ^{13}C HSQC analysis contained 0.25% $^{13}\text{C}_6$ -glucose (Cambridge Isotope Laboratories). For 2D ^1H , ^{13}C HSQC and 2D ^1H , ^1H TOCSY involving ethanol stress, deuterated ethanol (Isotec) was used to minimize the contribution of exogenous ethanol to the NMR spectra. For 1D ^1H NMR experiments, 2.74 OD₆₀₀ units were harvested at each time point and for the 2D ^1H , ^{13}C HSQC and 2D ^1H , ^1H TOCSY experiments, 5.48 OD₆₀₀ units were collected. Bacteria were harvested by centrifugation (4000 rpm for 5 min), suspended in 50 mM phosphate buffer in 100% D₂O 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 $-80\text{ }^{\circ}\text{C}$ until ready for analysis. At the time of use, a 600 L 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 1D ^1H 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 (\vec{P}_1) corresponding to the largest variation in the data set within this multidimensional space. The second vector (\vec{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 data set's variability, 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 \vec{P}_1 and \vec{P}_2 . The PC1 and PC2 scores are usually presented in a 2D 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 2D scores plot.³¹ Each NMR spectrum was center averaged for PCA analysis to minimize any experimental variations between cultures.³²

2D ^1H , ^{13}C HSQC spectra were collected and processed as previously described.⁷ 2D ^1H , ^1H TOCSY spectra were collected with watergate solvent presaturation,³³ and a relaxation delay of 2 seconds. A total of 1024 data points with a sweep width of 5000 Hz, and 256 data points with a sweep width of 5001.324 Hz were collected in the direct and indirect ^1H dimensions, respectively. A total of 16 dummy scans and 8 acquisition scans were used to obtain each of the 2D ^1H , ^1H TOCSY NMR spectra. The 2D ^1H , ^1H TOCSY NMR spectra were processed similar to the 2D ^1H , ^{13}C HSQC spectra and both spectra were analyzed using NMRView³⁴ (One Moon Scientific) and Sparky (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco) to identify chemical shifts and assign peak intensities.

The observed NMR peaks in the 2D ^1H , ^{13}C HSQC and ^1H , ^1H TOCSY spectra were assigned to specific metabolites using ^1H and ^{13}C chemical shift tolerances of 0.05 ppm and 0.50 ppm, respectively, and the Madison Metabolomics Consortium Database (MMCD),³⁵ the BioMagResBank,³⁶ and the Human Metabolome Database.³⁷ The presence of metabolites and metabolic pathways were verified with the KEGG³⁸ and Metacyc³⁹ databases.

Peak intensities were normalized for each 2D 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 2D spectra for strain 1457, 1457-*acnA*, ethanol stress and iron-limitation. 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 relative to the wild-type bacteria in TSB medium. Peaks with calculated percent differences greater than 5 times the average percent 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 in order 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 2D scores plots were quantitatively analyzed using a tree diagram and bootstrapping technique.⁴⁰ The PC1 and PC2 scores for each set of 10 duplicate NMR spectra representing a specific metabolic state (iron-limitation, ethanol treatment, *etc.*) are used to calculate an average PC score and standard deviation. Any PC score 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 2D 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 2D scores plot, standard bootstrapping methods

were also applied.^{41,42} Briefly, the average PC scores are recalculated by randomly selecting points from the data set. Distances are recalculated between the clusters using the new average PC scores to create a new distance matrix. The process is repeated until 100 different distance matrices are created and transferred to version 3.68 of the PHYLIP⁴³ suite of software programs (<http://www.phylip.com>). 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 if ethanol stress and iron-limitation alter TCA cycle activity, the specific activity of aconitase in *S. epidermidis* strain 1457 at 2 hours (exponential growth) and 6 hours (post-exponential growth) post-inoculation was assessed (Figure 6.1). As 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 post-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.

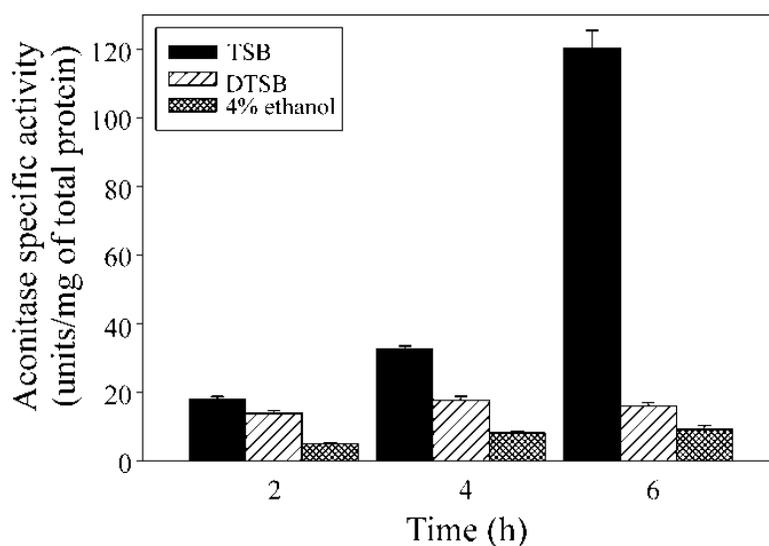


Figure 6.1 The temporal induction of aconitase specific activity is inhibited by dissimilar stressors. Aconitase activity was assessed during the exponential (2 and 4 h) and post-exponential (6 h) growth phases during growth in DTSB or TSB containing 4% ethanol. The data are presented as the mean and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

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 two or six hours in TSB, TSB with 4% ethanol, or DTSB. Following acquisition of the NMR spectra, the σ of integrals was used for PCA (Figure 6.3a). 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 6.2). This was expected due to the normal repression of TCA cycle activity during nutrient rich growth (Figure 6.1^{8,44}). 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 6.3a 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 aconitase mutant strains, suggest that ethanol stress had TCA cycle-independent metabolomic effects (Figure 6.3a). 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.

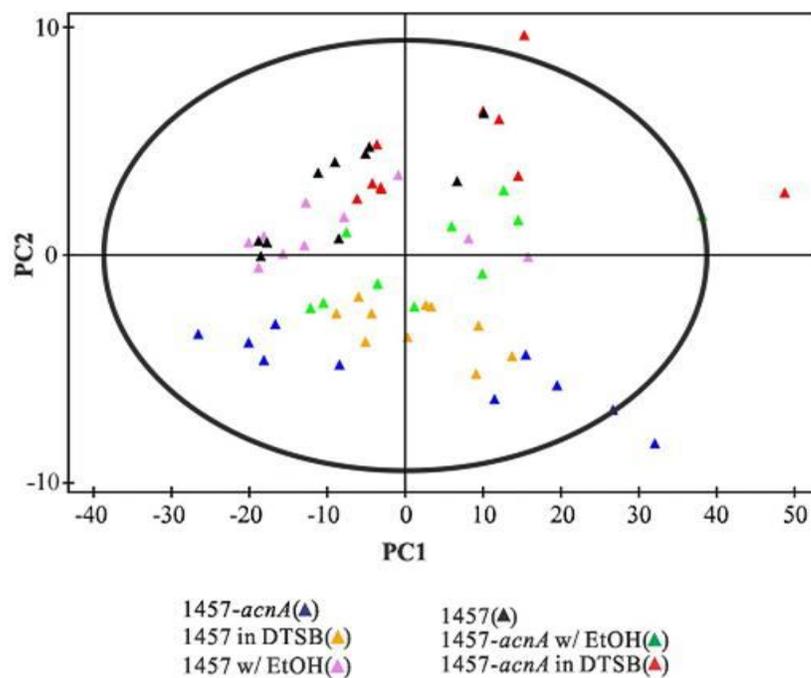


Figure 6.2 PCA 2D scores plot comparing non-stressed, or iron-limited cultures of strains 1457, 1457-*acnA*, and 1457-*sigB::dhfr* grown for 2h. Symbols and colors are defined in the figure.

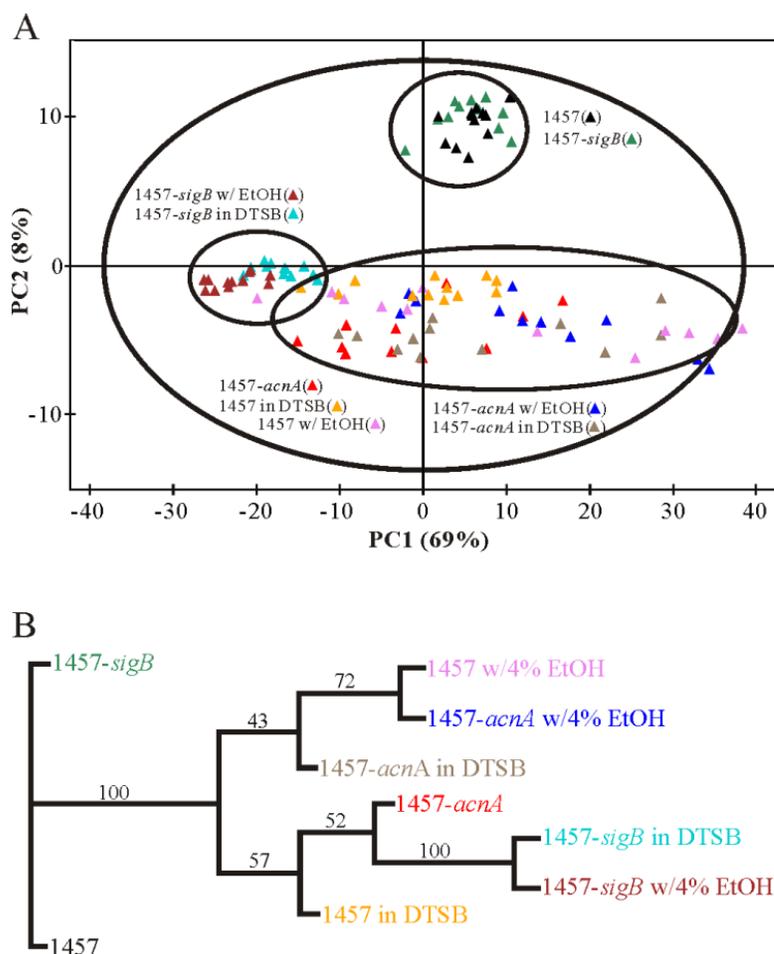


Figure 6.3 Environmental stressors cause metabolomic perturbations similar to TCA cycle inactivation. (A) PCA 2D scores plot comparing non-stressed, ethanol stressed, or iron-limited cultures of strains 1457, 1457-*acnA*, and 1457-*sigB*::*dhfr* grown for 6 h. Symbols and colors are defined in the figure. 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 parenthesis. (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.

The common metabolomic response to environmental stimuli can be more easily observed by a recently developed method to visualize PCA data.⁴⁰ 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 (Figure 6.3b).⁴³ Using this approach, it becomes clear that stress induced metabolomic responses are very similar to the metabolome of the aconitase deficient strain 1457-*acnA*. As with the 2D scores plot (Figure 6.3a), the higher bootstrap values in the dendrogram for ethanol stressed cultures also indicate that ethanol has TCA cycle-independent effects on the metabolome.

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

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

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

^b Due to peak overlap, we are unable to determine if the metabolite is acetyl-glutamine or acetyl-glutamate; however, we note that acetyl-glutamine is uncommon in prokaryotes. Metabolites whose concentrations were changed under all stress conditions relative to the wild-type strain grown in TSB are shaded in grey.

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

Strain 1457 ^a (DTSB) Iron-limitation	Strain 1457 ^a (4% Ethanol) Ethanol stress	Strain 1457- <i>acnA</i> ^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		
<i>N</i> -Acetyl-glucosamine	<i>N</i> -Acetyl-glucosamine Acetyl-glucosamine 6-phosphate	
<i>N</i> -Acetyl-mannosamine	<i>N</i> -Acetyl-mannosamine	<i>N</i> -Acetyl-mannosamine
<i>N</i> -Acetyl-neuraminic acid	<i>N</i> -Acetyl-neuraminic acid	<i>N</i> -Acetyl-neuraminic acid
Acetyl-phosphate	Acetyl-phosphate	Acetyl-phosphate
		Arginine
		Citrate
Ethanol		
Glucosamine	Glucosamine ^c	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- <i>N</i> -acetyl-glucosamine	
Metabolites whose concentration is decreased relative to strain 1457 grown in TSB medium.		
α -Ketoglutarate	α -Ketoglutarate	α -Ketoglutarate
γ -Aminobutyrate	γ -Aminobutyrate	γ -Aminobutyrate
Acetyl-glutamate ^b	Acetyl-glutamate ^b	Acetyl-glutamate ^b
		Acetyl-ornithine
		Alanine
Arginine		
Asparagine	Asparagine	Asparagine
		Aspartate
β -Alanine	β -Alanine	β -Alanine

Citrulline	Citrulline	Citrulline
	Ethanol	Ethanol
Fructose-6-Phosphate	Fructose-6-Phosphate	Fructose-6-phosphate
Glutamate	Glutamate	Glutamate
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- <i>N</i> -acetyl-glucosamine		

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

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

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

Metabolites whose concentrations were changed under all stress conditions relative to the wild-type strain grown in TSB are shaded in grey.

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.^{8,44} 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 un-stressed exponential growth phase (2 h) metabolome than to the unstressed metabolome of post-exponential growth phase of cultures. As expected, PCA of unstressed strain 1457 cultures, grown for 2 or 6 hours, form separate subsets in a 3D scores plot (Figure 6.4). Consistent with our hypothesis, PCA of post-exponential growth phase stressed and *acnA* mutant cultures were more closely 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 which interferes with TCA cycle function results in a metabolome similar to an unstressed exponential phase culture.

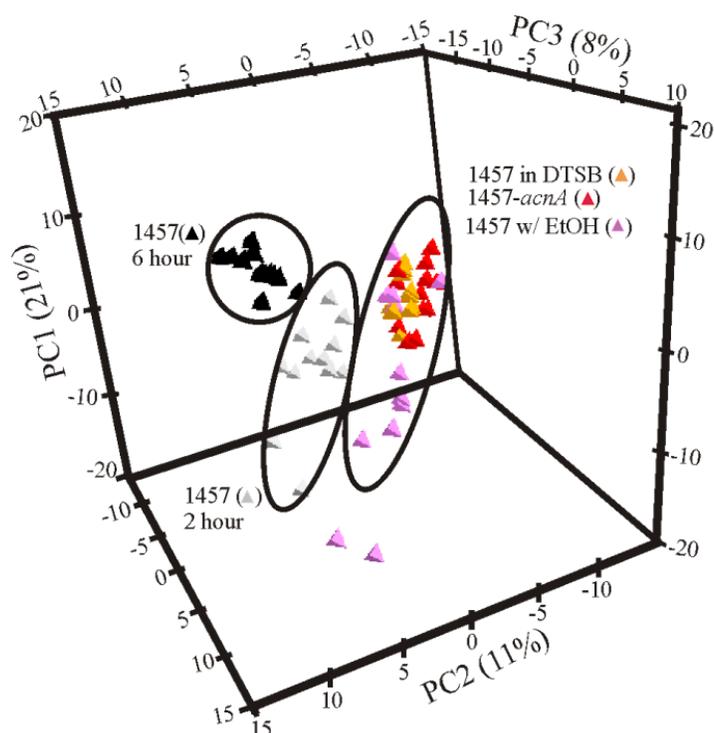


Figure 6.4 PCA 3D scores plot comparing non-stressed, ethanol stressed, or iron-limited cultures of strains 1457 grown for 6 h to that of strain 1457 grown for 2 h. Symbols and colors are defined in the figure. The ovals are manually drawn to identify clusters of related samples and to guide the reader. They are not statistically relevant. The relative contribution of each principal component is indicated within the parenthesis.

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.^{45,46} 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.⁴⁷⁻⁴⁹ 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 two or six hours 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 metabolomic changes associated with iron-limitation and ethanol stress occur largely independent of σ^B (Figure 6.3a). Additionally, the metabolomic dendrogram confirms 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 6.3b). 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: 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*; a.k.a. *glnA*) (Figure 6.5). The data suggest the first possibility is the more correct one; specifically, bacteria are responding to metabolomic 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 2 and 5 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). 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 post-exponential growth phases. The correlation between Gln and Glu concentrations and *glnA* mRNA levels are 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.

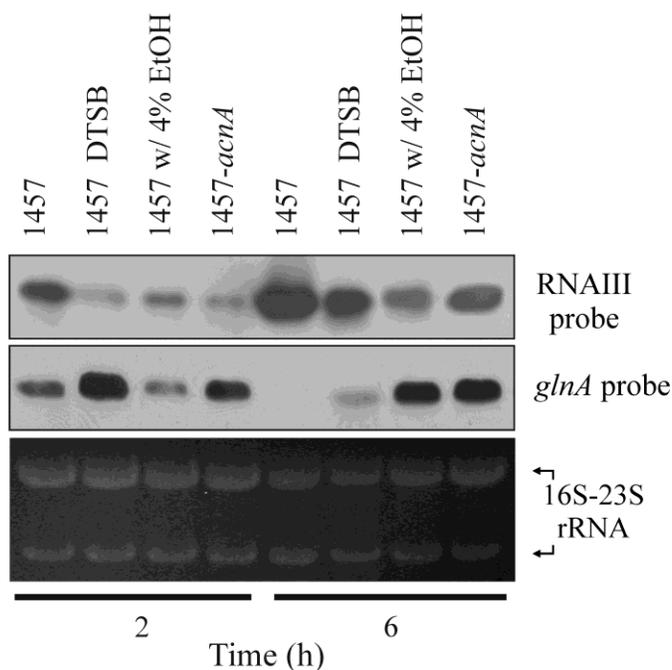


Figure 6.5 Northern blot analysis of RNAIII and *glnA* mRNA levels in the exponential (2 h) and post-exponential (6 h) phases of growth. To ensure that equivalent quantities of RNA were loaded in the gel, 23S and 16S rRNA were visualized by ethidium bromide staining and used as loading controls (bottom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

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.⁵⁰ The increase in phosphorylated HPr enhances its interaction with the catabolite control protein A (CcpA).⁵¹⁻⁵³ CcpA primarily functions as

a repressor; however, it also activates transcription of genes involved in fermentation and overflow metabolism.^{52,54} In addition to the concentration of glucose-6-phosphate being increased by TCA cycle stress, the concentration of several fermentation products or intermediates (*i.e.*, lactate, acetate, and acetaldehyde) and the small phosphodonor 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;⁵⁴ however, CodY responds to the intracellular concentrations of branched chain amino acids (BCAA).⁵⁵ TCA cycle stress did not alter the concentrations of BCAA beyond the 5-fold threshold (Table 6.1), suggesting the increase in overflow metabolism was independent of CodY.

In *S. aureus*, CcpA enhances biofilm formation and PIA biosynthesis, while CodY represses PIA synthesis.^{56,57} In *S. epidermidis*, ethanol stress, iron-limitation, and TCA cycle inactivation enhance biofilm formation and PIA synthesis.^{7,13,14,16,24,58} 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 hours of growth (Figure 6.6). Consistent with previous observations,⁷ TCA cycle inactivation (strain 1457-*acnA*) dramatically increased the accumulation of PIA, while 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 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 to produce PIA, strongly suggesting that some TCA cycle-associated metabolomic changes (*i.e.*, glucose-6-phosphate) are “sensed” by CcpA, which in turn activates PIA biosynthesis.

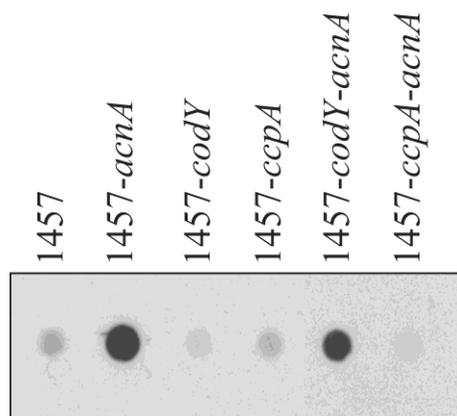


Figure 6.6 CcpA is required for PIA synthesis during TCA cycle stress. PIA immunoblot assay of strain 1457 and isogenic mutants of *acnA*, *codY*, *ccpA*, *codY/acnA*, and *ccpA/acnA* grown for 6 h in TSB. The results are representative of 3 independent experiments.

6.3.6 TCA cycle stress decreases RNAIII transcription

In *S. aureus*, inactivation of the TCA cycle increases the transcription or stability of the accessory gene regulator (Agr) system’s riboregulator RNAIII.^{8,9} Conversely, increasing TCA cycle activity decreases the transcription or stability of RNAIII.⁵⁹ The causal relationship between TCA cycle activity and RNAIII transcript levels in *S. aureus* led us to examine if disparate environmental conditions would similarly affect RNAIII

transcription or message stability in *S. epidermidis*. In contrast to *S. aureus*, TCA cycle inactivation decreased RNAlIIII transcription or stability during the exponential and post-exponential growth phases in *S. epidermidis* (Figure 6.5). Importantly, ethanol stress and iron-limitation decreased RNAlIIII 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 RNAlIIII).

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 transcription of virulence genes; thus, environmental signals (*e.g.*, nutrient replete, iron-limiting, or oxygen-limiting growth conditions) commonly regulate virulence gene transcription.^{2,60-64} Although *S. epidermidis* has relatively few virulence determinants, one of its primary pathogenic effectors is the exopolysaccharide PIA.⁶⁵⁻⁶⁸ 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.^{7,24,59} In this study, we demonstrate that dissimilar environmental signals decrease TCA cycle activity (Figure 6.1) resulting in common metabolomic changes (Figure 6.3 and Tables 6.1 and 6.2; summarized in Figure 6.7) that alter the activity of metabolite-responsive regulators such as CcpA (Figure 6.7).

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.

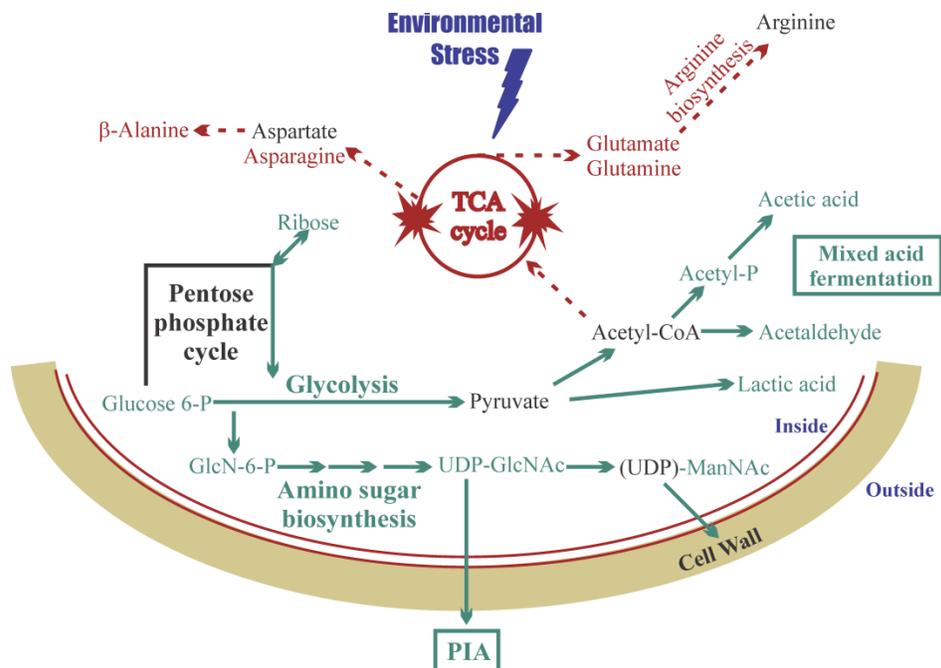


Figure 6.7 Summary of post-exponential growth phase metabolic changes associated with TCA cycle stress. Metabolites in green represent an increased concentration relative to the wild-type strain. Metabolites in red represent a decreased concentration relative to the wild-type strain. Metabolites and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

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} M⁶¹, 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 O₂ is 159 mmHg at sea level and this decreases to an estimated 3-5 mmHg 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 five-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^{13,69}). Third, these data suggest a difficulty in attributing the effects of an environmental stimulus, such as iron-limitation, to a specific regulator like the ferric uptake regulator (Fur),⁷⁰ 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 6.3). Finally, in bacteria, three metabolic pathways (Embden-Meyerhof-Parnas, pentose phosphate, and TCA cycle) produce the thirteen 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 6.2) 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 metabolomic 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.; Peters, G.; Heilmann, C. *Lancet Infect Dis* **2002**, 2, 677.
- (2) Somerville, G. A.; Proctor, R. A. *Microbiol Mol Biol Rev* **2009**, 73, 233.
- (3) Begun, J.; Sifri, C. D.; Goldman, S.; Calderwood, S. B.; Ausubel, F. M. *Infect Immun* **2005**, 73, 872.

- (4) Coulter, S. N.; Schwan, W. R.; Ng, E. Y.; Langhorne, M. H.; Ritchie, H. D.; Westbrook-Wadman, S.; Hufnagle, W. O.; Folger, K. R.; Bayer, A. S.; Stover, C. K. *Mol Microbiol* **1998**, *30*, 393.
- (5) Mei, J. M.; Nourbakhsh, F.; Ford, C. W.; Holden, D. W. *Mol Microbiol* **1997**, *26*, 399.
- (6) Bae, T.; Banger, A. K.; Wallace, A.; Glass, E. M.; Aslund, F.; Schneewind, O.; Missiakas, D. M. *Proc Natl Acad Sci U S A* **2004**, *101*, 12312.
- (7) Sadykov, M. R.; Olson, M. E.; Halouska, S.; Zhu, Y.; Fey, P. D.; Powers, R.; Somerville, G. A. *J Bacteriol* **2008**, *190*, 7621.
- (8) Somerville, G. A.; Chaussee, M. S.; Morgan, C. I.; Fitzgerald, J. R.; Dorward, D. W.; Reitzer, L. J.; Musser, J. M. *Infect Immun* **2002**, *70*, 6373.
- (9) Somerville, G. A.; Cockayne, A.; Dürr, M.; Peschel, A.; Otto, M.; Musser, J. M. *J Bacteriol* **2003**, *185*, 6686.
- (10) Collins, F. M.; Lascelles, J. J. *Gen. Microbiol.* **1962**, *29*, 531.
- (11) Somerville, G.; Mikoryak, C. A.; Reitzer, L. *J Bacteriol* **1999**, *181*, 1072.
- (12) Varghese, S.; Tang, Y.; Imlay, J. A. *J Bacteriol* **2003**, *185*, 221.
- (13) Knobloch, J. K.; Bartscht, K.; Sabottke, A.; Rohde, H.; Feucht, H. H.; Mack, D. *J Bacteriol* **2001**, *183*, 2624.
- (14) Presterl, E.; Suchomel, M.; Eder, M.; Reichmann, S.; Lassnigg, A.; Graninger, W.; Rotter, M. *J Antimicrob Chemother* **2007**, *60*, 417.
- (15) Korem, M.; Gov, Y.; Shirron, N.; Shuster, A.; Rosenberg, M. *FEMS Microbiol Lett* **2007**, *269*, 153.

- (16) Lyte, M.; Freestone, P. P.; Neal, C. P.; Olson, B. A.; Haigh, R. D.; Bayston, R.; Williams, P. H. *Lancet* **2003**, *361*, 130.
- (17) Otzen, D. E.; Sehgal, P.; Nesgaard, L. W. *Biochemistry* **2007**, *46*, 4348.
- (18) Mack, D.; Siemssen, N.; Laufs, R. *Infect Immun* **1992**, *60*, 2048.
- (19) Handke, L. D.; Slater, S. R.; Conlon, K. M.; O'Donnell, S. T.; Olson, M. E.; Bryant, K. A.; Rupp, M. E.; O'Gara, J. P.; Fey, P. D. *Can J Microbiol* **2007**, *53*, 82.
- (20) Horton, R. M.; Cai, Z. L.; Ho, S. N.; Pease, L. R. *Biotechniques* **1990**, *8*, 528.
- (21) Kennedy, M. C.; Emptage, M. H.; Dreyer, J. L.; Beinert, H. *J Biol Chem* **1983**, *258*, 11098.
- (22) Baughn, A. D.; Malamy, M. H. *Proc Natl Acad Sci U S A* **2002**, *99*, 4662.
- (23) Lowry, O. H.; Rosebrough, N. J.; Farr, L.; Randall, R. J. *J Biol Chem* **1951**, *193*, 267.
- (24) Vuong, C.; Kidder, J. B.; Jacobson, E. R.; Otto, M.; Proctor, R. A.; Somerville, G. A. *J Bacteriol* **2005**, *187*, 2967.
- (25) Kay, L. E.; Keifer, P.; Saarinen, T. *J. Am. Chem. Soc.* **1992**, *114*, 10663.
- (26) Palmer III, A. G.; Cavanagh, J.; Wright, P. E.; Rance, M. *J. Magn. Reson.* **1991**, *93*, 151.
- (27) Bax, A.; Davis, D. G. *J. Magn. Reson.* **1985**, *65*, 355.
- (28) Forgue, P.; Halouska, S.; Werth, M.; Xu, K.; Harris, S.; Powers, R. *J Proteome Res* **2006**, *5*, 1916.

- (29) Halouska, S.; Chacon, O.; Fenton, R. J.; Zinniel, D. K.; Barletta, R. G.; Powers, R. *J Proteome Res* **2007**, *6*, 4608.
- (30) Halouska, S.; Powers, R. *J Magn Reson* **2006**, *178*, 88.
- (31) Nguyen, B. D.; Meng, X.; Donovan, K. J.; Shaka, A. J. *J Magn Reson* **2007**, *184*, 263.
- (32) Craig, A.; Cloarec, O.; Holmes, E.; Nicholson, J. K.; Lindon, J. C. *Anal Chem* **2006**, *78*, 2262.
- (33) Piotto, M.; Saudek, V.; Sklenar, V. *J Biomol NMR* **1992**, *2*, 661.
- (34) Johnson, B. A. *Methods Mol Biol* **2004**, *278*, 313.
- (35) Cui, Q.; Lewis, I. A.; Hegeman, A. D.; Anderson, M. E.; Li, J.; Schulte, C. F.; Westler, W. M.; Eghbalnia, H. R.; Sussman, M. R.; Markley, J. L. *Nat Biotechnol* **2008**, *26*, 162.
- (36) Ulrich, E. L.; Akutsu, H.; Doreleijers, J. F.; Harano, Y.; Ioannidis, Y. E.; Lin, J.; Livny, M.; Mading, S.; Maziuk, D.; Miller, Z.; Nakatani, E.; Schulte, C. F.; Tolmie, D. E.; Kent Wenger, R.; Yao, H.; Markley, J. L. *Nucleic Acids Res* **2008**, *36*, D402.
- (37) Wishart, D. S.; Tzur, D.; Knox, C.; Eisner, R.; Guo, A. C.; Young, N.; Cheng, D.; Jewell, K.; Arndt, D.; Sawhney, S.; Fung, C.; Nikolai, L.; Lewis, M.; Coutouly, M. A.; Forsythe, I.; Tang, P.; Shrivastava, S.; Jeroncic, K.; Stothard, P.; Amegbey, G.; Block, D.; Hau, D. D.; Wagner, J.; Miniaci, J.; Clements, M.; Gebremedhin, M.; Guo, N.; Zhang, Y.; Duggan, G. E.; Macinnis, G. D.; Weljie, A. M.; Dowlatabadi, R.; Bamforth, F.; Clive, D.; Greiner, R.; Li, L.; Marrie, T.; Sykes, B. D.; Vogel, H. J.; Querengesser, L. *Nucleic Acids Res* **2007**, *35*, D521.
- (38) Kanehisa, M.; Araki, M.; Goto, S.; Hattori, M.; Hirakawa, M.; Itoh, M.; Katayama, T.; Kawashima, S.; Okuda, S.; Tokimatsu, T.; Yamanishi, Y. *Nucleic Acids Res* **2008**, *36*, D480.

- (39) Karp, P. D.; Ouzounis, C. A.; Moore-Kochlacs, C.; Goldovsky, L.; Kaipa, P.; Ahren, D.; Tsoka, S.; Darzentas, N.; Kunin, V.; Lopez-Bigas, N. *Nucleic Acids Res* **2005**, *33*, 6083.
- (40) Werth, M. T.; Halouska, S.; Shortridge, M. D.; Zhang, B.; Powers, R. *Anal Biochem* **2009**.
- (41) Efron, B.; Halloran, E.; Holmes, S. *Proc Natl Acad Sci U S A* **1996**, *93*, 13429.
- (42) Felsenstein, J. *Evolution* **1985**, *39*, 783.
- (43) Retief, J. D. *Methods Mol Biol* **2000**, *132*, 243.
- (44) Somerville, G. A.; Sa ïl-Salim, B.; Wickman, J. M.; Raffel, S. J.; Kreiswirth, B. N.; Musser, J. M. *Infect Immun* **2003**, *71*, 4724.
- (45) Wu, S.; de Lencastre, H.; Tomasz, A. *J Bacteriol* **1996**, *178*, 6036.
- (46) Bischoff, M.; Dunman, P.; Kormanec, J.; Macapagal, D.; Murphy, E.; Mounts, W.; Berger-Bachi, B.; Projan, S. *J Bacteriol* **2004**, *186*, 4085.
- (47) Bischoff, M.; Entenza, J. M.; Giachino, P. *J Bacteriol* **2001**, *183*, 5171.
- (48) Meier, S.; Goerke, C.; Wolz, C.; Seidl, K.; Homerova, D.; Schulthess, B.; Kormanec, J.; Berger-Bachi, B.; Bischoff, M. *Infect Immun* **2007**, *75*, 4562.
- (49) Pane-Farre, J.; Jonas, B.; Forstner, K.; Engelmann, S.; Hecker, M. *Int J Med Microbiol* **2006**, *296*, 237.
- (50) Deutscher, J.; Saier, M. H., Jr. *Proc Natl Acad Sci U S A* **1983**, *80*, 6790.
- (51) Bruckner, R.; Titgemeyer, F. *FEMS Microbiol Lett* **2002**, *209*, 141.

- (52) Sonenshein, A. L. *Nat Rev Microbiol* **2007**, *5*, 917.
- (53) Warner, J. B.; Lolkema, J. S. *Microbiol Mol Biol Rev* **2003**, *67*, 475.
- (54) Shivers, R. P.; Dineen, S. S.; Sonenshein, A. L. *Mol Microbiol* **2006**, *62*, 811.
- (55) Shivers, R. P.; Sonenshein, A. L. *Mol Microbiol* **2004**, *53*, 599.
- (56) Seidl, K.; Goerke, C.; Wolz, C.; Mack, D.; Berger-Bachi, B.; Bischoff, M. *Infect Immun* **2008**, *76*, 2044.
- (57) Majerczyk, C. D.; Sadykov, M. R.; Luong, T. T.; Lee, C.; Somerville, G. A.; Sonenshein, A. L. *J Bacteriol* **2008**, *190*, 2257.
- (58) Matinaho, S.; von Bonsdorff, L.; Rouhiainen, A.; Lonroth, M.; Parkkinen, J. *FEMS Microbiol Lett* **2001**, *196*, 177.
- (59) Zhu, Y.; Xiong, Y. Q.; Sadykov, M. R.; Fey, P. D.; Lei, M. G.; Lee, C. Y.; Bayer, A. S.; Somerville, G. A. *Infect Immun* **2009**, *77*, 4256.
- (60) Bullen, J. J.; Rogers, H. J.; Griffiths, E. *Curr Top Microbiol Immunol* **1978**, *80*, 1.
- (61) Litwin, C. M.; Calderwood, S. B. *Clin Microbiol Rev* **1993**, *6*, 137.
- (62) Mekalanos, J. J. *J Bacteriol* **1992**, *174*, 1.
- (63) Park, M. K.; Myers, R. A.; Marzella, L. *Clin Infect Dis* **1992**, *14*, 720.
- (64) Milenbachs, A. A.; Brown, D. P.; Moors, M.; Youngman, P. *Mol Microbiol* **1997**, *23*, 1075.

- (65) Vuong, C.; Kocianova, S.; Voyich, J. M.; Yao, Y.; Fischer, E. R.; Deleo, F. R.; Otto, M. *J Biol Chem* **2004**, *24*, 54881.
- (66) Vuong, C.; Voyich, J. M.; Fischer, E. R.; Braughton, K. R.; Whitney, A. R.; DeLeo, F. R.; Otto, M. *Cell Microbiol* **2004**, *6*, 269.
- (67) Rupp, M. E.; Ulphani, J. S.; Fey, P. D.; Mack, D. *Infect Immun* **1999**, *67*, 2656.
- (68) Rupp, M. E. In *The Staphylococci in Human Disease*; K, C., G, A., Eds.; Churchill-Livingstone: New York, 1997, p 379.
- (69) Deighton, M.; Borland, R. *Infect Immun* **1993**, *61*, 4473.
- (70) Torres, V. J.; Attia, A. S.; Mason, W. J.; Hood, M. I.; Corbin, B. D.; Beasley, F. C.; Anderson, K. L.; Stauff, D. L.; McDonald, W. H.; Zimmerman, L. J.; Friedman, D. B.; Heinrichs, D. E.; Dunman, P. M.; Skaar, E. P. *Infect Immun* **2010**, *78*, 1618.

CHAPTER 7

THE INFLUENCE OF IRON AND AERATION ON *STAPHYLOCOCCUS AUREUS* GROWTH, METABOLISM, AND TRANSCRIPTION

7.1 Introduction

Staphylococcus aureus is a versatile pathogen capable of infecting almost any niche within a human or animal host. As with most bacterial pathogens, successful pathogenesis requires that *S. aureus* gain entry into a host (*e.g.*, through a breach in the skin), adhere to a suitable surface, avoid being killed by host's immune system, acquire nutrients, and proliferate. Notably, every step in this pathogenic process involves metabolic changes and/or metabolic regulation. Entry into a host is an environmental change that dramatically alters the availability of nutrients and co-factors. Similarly, the transition from synthesizing adhesins to producing tissue-damaging secreted virulence determinants is preceded by changes in the bacterial nutritional status.¹ To avoid being killed by the host's immune system, *S. aureus* alters its metabolism to limit damage by oxidative and nitrosative stress.² The act of acquiring nutrients is regulated by metabolite-responsive regulators that respond to changes in the bacterial metabolic status.³ Finally, proliferation is a carbon and energy intensive undertaking. In short, *S. aureus* success as a pathogen depends on its ability to meet the rapidly changing nutritional and energy requirements necessary for survival and proliferation within a host. The ability to meet the changing nutritional and energy requirements is complicated by the host, which limits

access to enzymatic co-factors such as oxygen and iron by sequestering these in host proteins and molecules (*e.g.*, hemoglobin, heme, and transferrin). The limited availability of these co-factors creates metabolic blocks that hinder virulence factor synthesis,^{4,5} which leaves *S. aureus* vulnerable to the host's immune system.

The potential consequences of a metabolic block(s) (*i.e.*, decreased fecundity, fitness, and/or death) create selective pressure to maintain a functional metabolic state. Over time, this selective pressure has led to the evolution/acquisition of metabolite-responsive regulators (*e.g.*, CodY, CcpA, RpiRc) that maintain metabolic homeostasis when nutritional and environmental conditions are favorable, and that facilitate metabolic adaptations when nutritional and environmental conditions change.⁶⁻⁸ Specifically, CodY responds to changes in branched chain amino acids and GTP concentrations,^{9,10} while CcpA indirectly responds to glucose-6-phosphate and fructose-1,6-bisphosphate.¹¹ While the RpiR regulators all have sugar isomerases binding domains, the metabolites that modulate their activities have yet to be determined.⁸ These metabolite-responsive regulators, *i.e.* CodY, CcpA, and RpiRc, coordinate both virulence determinant biosynthesis and metabolism in staphylococci.^{8,9,12} In general, the activity of metabolite-responsive regulators is controlled by intracellular concentrations of numerous compounds, such as biosynthetic intermediates,¹³ amino acids,¹⁴ nucleic acids,¹⁵ and co-factors (*e.g.*, iron).¹⁶ In *Staphylococcus epidermidis*, the intracellular concentrations of metabolites can be altered by environmental stressors, such as antibiotics, ethanol, metal ion-limitation, or nutrient-depletion.^{3,17,18} This creates a mechanism by which extracellular stress can be transduced into intracellular metabolic signals that elicit

regulatory changes via metabolite-responsive regulators. While the transduction of external factors into internal metabolic signals has been demonstrated in *S. epidermidis*, it has not been shown in *S. aureus*.

Two important enzymatic co-factors that influence *S. aureus* fitness and that are controlled by the host are oxygen and iron. Numerous studies have shown that oxygen- and iron-availability can influence growth, antibiotic tolerance, and synthesis of virulence factors;¹⁹⁻²³ however, little regard has been given to the metabolic changes that accompany iron- and oxygen-limitation.²⁴⁻²⁶ Where the metabolic effects of iron- or oxygen-limitation have been examined, those studies only addressed the metabolic effects of a single stress and in a single growth phase/state. The analysis of a single growth phase/state overlooks the fact that staphylococci have two very distinct metabolic states that correlate with the exponential and post-exponential growth phases.²⁷ Because pathogenic bacteria rarely encounter a single stress, we chose to examine the combined stresses of iron- and oxygen-limitation on the exponential and post-exponential metabolomes. In addition, we also assessed the contributions of the single stresses (*i.e.*, iron- or oxygen-limitation) to the overall effects of combined stresses (*i.e.*, iron- and oxygen-limitation). Lastly, by analyzing the metabolic effects of single stresses, we can identify common metabolic signals, similar to what was done in *S. epidermidis*.^{17,18} In the current study, the effects of oxygen- and iron-limitation on the metabolome of *S. aureus* strain SA564 were examined using NMR metabolomics. In addition, a select set of genes were examined for transcriptional changes that might be associated with the altered growth conditions.

7.2 Experimental procedures

7.2.1 Bacterial strains, media, and cultivation conditions

S. aureus strain SA564 was grown in tryptic soy broth (TSB; BD Biosciences) or on tryptic soy agar (TSA). For iron-limitation cultivation conditions, deferrated TSB (DTSB) medium was used.¹⁷ In this study, iron-limited medium was preferred to “iron-free” medium because iron-free medium requires the addition of adulterants, such as deferoxamine mesylate. These adulterants are added to chelate iron; however, they also chelate other cations,²⁸ which makes comparison between cultivation conditions impractical because more than one variable would be changed. To prevent pyrolysis of labile metabolites, DTSB and TSB were prepared by filter sterilization. For iron-limiting conditions, bacterial cultures were grown overnight in 10 mL of DTSB, harvested by centrifugation at 3,795 x g, washed once in 10 mL of DTSB, and suspended in fresh DTSB. This culture was used to inoculate (1:100) the starter cultures, which were cultivated in 100 mL of DTSB or TSB for 2 h. After 2 h of growth, these cultures were used to inoculate the primary cultures to an optical density at 600 nm (O.D.₆₀₀) of 0.06. Each primary culture was cultivated in pre-warmed DTSB or TSB with defined flask-to-media ratios. All cultures were grown at 37 °C with constant aeration (225 rpm) and growth (O.D.₆₀₀) and pH were measured at hourly intervals for 12 h. For the purpose of varying culture aeration during growth, flask-to-media ratios (v/v) of 10:8, 10:4, 10:2.5, 10:1 were used. This approach takes advantage of the poor diffusion coefficient for oxygen into water to alter the amount of oxygen available to the bacteria.²⁹

7.2.2 Sample preparation for NMR metabolomic analysis

Bacterial cultures were harvested during the exponential (2 h) and post-exponential (6 h) growth phases. For one-dimensional (1D) ^1H NMR experiments, eight biological replicates (10 O.D.₆₀₀ units each) were harvested for each cultivation condition and growth phase. For two-dimensional (2D) ^1H , ^{13}C Heteronuclear Single Quantum Coherence (HSQC) experiments, four biological replicates (20 O.D.₆₀₀ units each) were harvested for each cultivation condition and growth phase. Bacteria were harvested using a sterile 0.45 μm Microfil V filtration system (EMD Millipore Corporation) that had been pre-washed with 5 mL of sterile phosphate buffered saline (PBS, pH 7.4). Following filtration, bacteria were washed twice with 5 mL of ice-cold PBS, each membrane was transferred to a 50 mL conical tube, and quenched in liquid nitrogen. After quenching, the conical tubes were placed in ice and bacteria were collected from the filter using ice-cold 20 mM phosphate buffer (1 mL). The bacterial suspensions were normalized to equal O.D.₆₀₀ units (10 for 1D NMR and 20 for 2D NMR) in a 1 mL total volume, transferred to Lysing matrix B tubes, and lysed twice (Speed 6 and 40 s) using a FastPrep FP120 instrument (MP Biomedicals). The samples were rested on ice for 5 min between lysings. To remove cell debris and glass beads, samples were centrifuged at 17,000 $\times g$ for 2 min at -9°C and 700 μL of supernatant from each tube was transferred into a new 2 mL microfuge tube (pre-cooled to -20°C). To the residual sample in the lysing matrix B tubes, 1 mL of ice-cold phosphate buffer was added, mixed, and centrifuged. After the second centrifugation, 900 μL of the supernatant from each tube were pooled with the samples from the first centrifugation. Of the 1.6 mL volume for each sample, 100 μL was

removed and the protein concentration was determined using a modified Lowry assay kit (Thermo Fisher Scientific). The remaining 1.5 mL of each sample was lyophilized and 600 μL of D_2O was added prior to analysis. 50 μM 3-(trimethylsilyl) propionic acid-2,2,3,3-d₄ (TMSP-d₄) or 500 μM TMSP-d₄ (Sigma-Aldrich) was added to each sample as internal standards for 1D ^1H NMR or 2D ^1H , ^{13}C HSQC experiments, respectively.

7.2.3 Data Collection

1D data collection was performed as described,³⁰ while 2D time zero extrapolated ^1H , ^{13}C HSQC (HSQC₀) NMR spectra were processed as described.³¹ For each sample, a total of 1024 data points with a spectrum width of 10.00 ppm in the ^1H dimension, and 64 data points with a spectrum width of 140.00 ppm in the ^{13}C dimension were collected. A total of 16 dummy scans and 64 scans with a receiver gain of 9195.2 and a relaxation delay of 1.5 seconds were applied for each sample.

7.2.4 Data analysis

1D ^1H NMR spectra were processed as described.¹⁸ 2D ^1H , ^{13}C HSQC spectra were processed using the NMRPipe software package.³² Peak picking and peak matching were done using NMRViewJ Version 8.³³ A table of peaks along with respective intensities were recorded and the concentrations of metabolites were calculated using a standard curve generated using nine metabolites with known concentrations: [D-glucose- $^{13}\text{C}_6$, D-fructose- $^{13}\text{C}_6$, glycine- $^{13}\text{C}_2$, DL-alanine-3- ^{13}C , sodium pyruvate- $^{13}\text{C}_3$, succinic acid- $^{13}\text{C}_4$, sodium acetate- $^{13}\text{C}_2$, 2-keto-3-(methyl- ^{13}C)-butyric acid-4- ^{13}C , sodium salt,

and 2-ketobutyric acid-4-¹³C sodium salt hydrate (Sigma-Aldrich)].³¹ A Student's T-test was used to determine the statistical significance ($p < 0.05$) of metabolite concentration changes.

The principal component analysis (PCA) and Orthogonal Projections to Latent Structures-discriminant analysis (OPLS-DA) scores plots and shared and unique structure (SUS)-plots were generated using the SIMCA 12.0+ (UMETRICS) statistical package (<http://www.umetrics.com/>). The OPLS-DA models were validated using a modified leave-one-out method^{34,35} and CV-ANOVA.³⁶ The metabolomics tree diagrams (dendrograms) and the ellipses corresponding to the 95% confidence limits from a normal distribution for each cluster within the PCA scores plots were generated using our PCA/PLS-DA utilities (<http://bionmr.unl.edu/pca-utils.php>).^{37,38}

7.2.5 Data interpretation

The peaks were assigned to metabolites using chemical shift references from the Human Metabolomics Database (HMDB),³⁹ Platform for RIKEN Metabolomics (PRIME)⁴⁰ Biological Magnetic Resonance Data Bank (BMRB)⁴¹ and Chenomx NMR Suite 7.6 (Chenomx. Inc., Edmonton, Canada) software. Heat maps were generated in R with a gplots package.⁴²

7.2.6 ICP-MS analysis of DTSB

Cultivation of *S. aureus* strain SA564 was performed as described. For the determination of iron concentrations in the culture supernatants, 1 mL of each culture was

harvested during the exponential (2 h) or post-exponential (6 h) growth phases by centrifugation at 3,795 x g for 5 min at room temperature. To remove any residual bacteria, samples were passed through Nalgene 0.2 µm polyethersulfone membrane syringe filters (Thermo Fisher Scientific). All samples were diluted 1:20 with a solution containing 52.6 ppb ⁷¹Ga in 0.1% HNO₃, which resulted in a ⁷¹Ga concentration of 50 ppb that was used as an internal standard. ICP-MS analyses were performed at the University of Nebraska-Lincoln Spectroscopy Core Facility using an Agilent Technologies ICP-MS 7500cs with an ESI SC-4 high-throughput autosampler (Elemental Scientific Inc.). The concentration of metals was calculated using ChemStation for ICP software with a calibration curve for 18 elements from serial dilutions of a standard stock from Inorganic Ventures. The instrument was operated with an octopole collision reaction cell filled with He at a flow rate of 5.0 mL/min and the following Ar plasma conditions: plasma power, 1500W, plasma gas, 17 L/min, auxiliary gas, 1 L/min., sample gas, 0.9-1.1 L/min and makeup gas 0.1-0.2 L/min, sample flow rate, 0.1 mL/min. For each sample, three replicates from three independent cultures were analyzed. Statistical significance was determined by one-way ANOVA using SigmaPlot 11.2 software (Systat Software Inc).

7.2.7 Aconitase activity assay

Aconitase enzymatic activity was measured as described.¹⁷ Protein concentrations were determined using a Modified Lowry Protein Assay Kit (Thermo Fisher Scientific).

7.2.8 Real-time RT-PCR

Total RNA was isolated from 10 O.D.₆₀₀ units of bacteria harvested in the exponential (2 h) and post-exponential (6 h) growth phases using the FastRNA Pro Blue kit (MP Biomedical) and the RNeasy kit (Qiagen) as described.⁴³ For mRNA analysis, a two-step real-time RT-PCR was performed with a cDNA synthesis from total RNA followed by PCR. Ten µg of TURBO DNase (Bio-Rad) treated total RNA for each sample was used to prepare cDNA by reverse transcriptase reaction using an iScript master mix (Bio-Rad) according to manufacturer's protocol. The reaction mixtures for Real-time PCR contained 10 µL of 2× SsoAdvanced SybrGreen Supermix (Bio-Rad), 7 pmoles of each primer (Table 1), and 5 µl of 10-fold diluted cDNA, in a total volume of 20 µl. As an internal reference, primers for 16s rRNA were used in the same reaction volume containing 1000-fold diluted cDNA as a template. The RT-PCR cycling conditions were: Initial denaturation at 94 °C for 3 min, followed by 40 cycles at 94 °C for 15 sec, 60 °C for 25 sec, and 72 °C for 20 sec. Relative transcript levels were determined by the comparative threshold ($\Delta\Delta C_T$) method (Bio-Rad). Experimental set-up and data analysis were carried out using CFX96 Real-Time PCR Detection System and Bio-rad CFX Manager Software version 3.0 (Bio-Rad).

Table 7.1 Real-time RT-PCR primers used in this study.

Primer	Mu50 orf*	Sequence
16SrRNA RTF	SAVrRNA16	CGTGCTACAATGGACAATACAAA
16SrRNA RTR	SAVrRNA16	ATCTACGATTACTAGCGATTCCA
<i>citB</i> RTF	SAV1350	GCGCAACAGCAACTGATTTA
<i>citB</i> RTR	SAV1350	GTTGTACACCTGGACCAAAGA
<i>feoB</i> RTF	SAV2584	GGAATGACAGCAACACAGTTAC
<i>feoB</i> RTR	SAV2584	GTGCTGACTGACCACCTAAA
<i>sbnA</i> RTF	SAV0116	TTCTGTAGGGCAAACACCTATG
<i>sbnA</i> RTR	SAV0116	GCTGCCTCCAGGATTCATATAC
<i>sstC</i> RTF	SAV0735	GACCTAATGGTGCGGGTAAG
<i>sstC</i> RTR	SAV0735	CAGACATGAGCTGTCCATCTATT

*Gene designations based on the *Staphylococcus aureus* strain Mu50 genomic DNA sequence⁴⁴

7.2.9 Acetate, glucose and lactate level determination of the media supernatant

Metabolite concentrations in the culture media were determined using kits purchased from R-biopharm, Inc.

7.3 Results

7.3.1 Growth of *S. aureus* under iron- and oxygen-limitation

The transition from an oxygen- and iron-replete environment (*e.g.*, the skin, the nares) to an oxygen- and iron-limited environment (*e.g.*, blood) constrains metabolic possibilities because many enzymatic reactions require iron and oxygen as co-factors. These constraints have fitness consequences; specifically, the ability to place progeny into the next generation.⁴⁵⁻⁴⁷ To assess the extent of the fitness cost of iron- and oxygen-limitation on the growth yield, *S. aureus* strain SA564 was cultivated under iron- and/or

oxygen-limited conditions and the growth was monitored (Figure 7.1a & 1b). As expected, the growth rates were largely independent of iron- and/or oxygen-limited conditions. Consistent with the Pasteur Effect, oxygen-limited growth conditions dramatically decreased biomass generation, whereas, iron-limitation did not significantly alter the biomass generated per mmol of glucose (Figure 7.1c).⁴⁸ Importantly, these data demonstrate the ease with which a batch culture can become microaerobic/anaerobic at an atmospheric oxygen concentration of 20.946% and with vigorous agitation at 225 rpm.²⁹ This transition to a microaerobic/anaerobic status is due to two factors: First, the poor diffusion coefficient for oxygen into water; and second, the relatively small surface area that is exposed to atmospheric oxygen as the volume of medium in a culture flask is increased. In contrast to oxygen-limitation, iron-limitation only decreased the growth yield of strain SA564 by ~25% relative to growth in iron-replete culture medium (Figure 7.1a). In *S. aureus*, the demand for iron is greatest in the post-exponential growth phase when the TCA cycle and the electron transport chain are most active; hence, the small amount of iron in the culture medium (Figure 7.1b) is sufficient to maintain the exponential growth rate but not the growth yield (Figure 7.1a).⁵

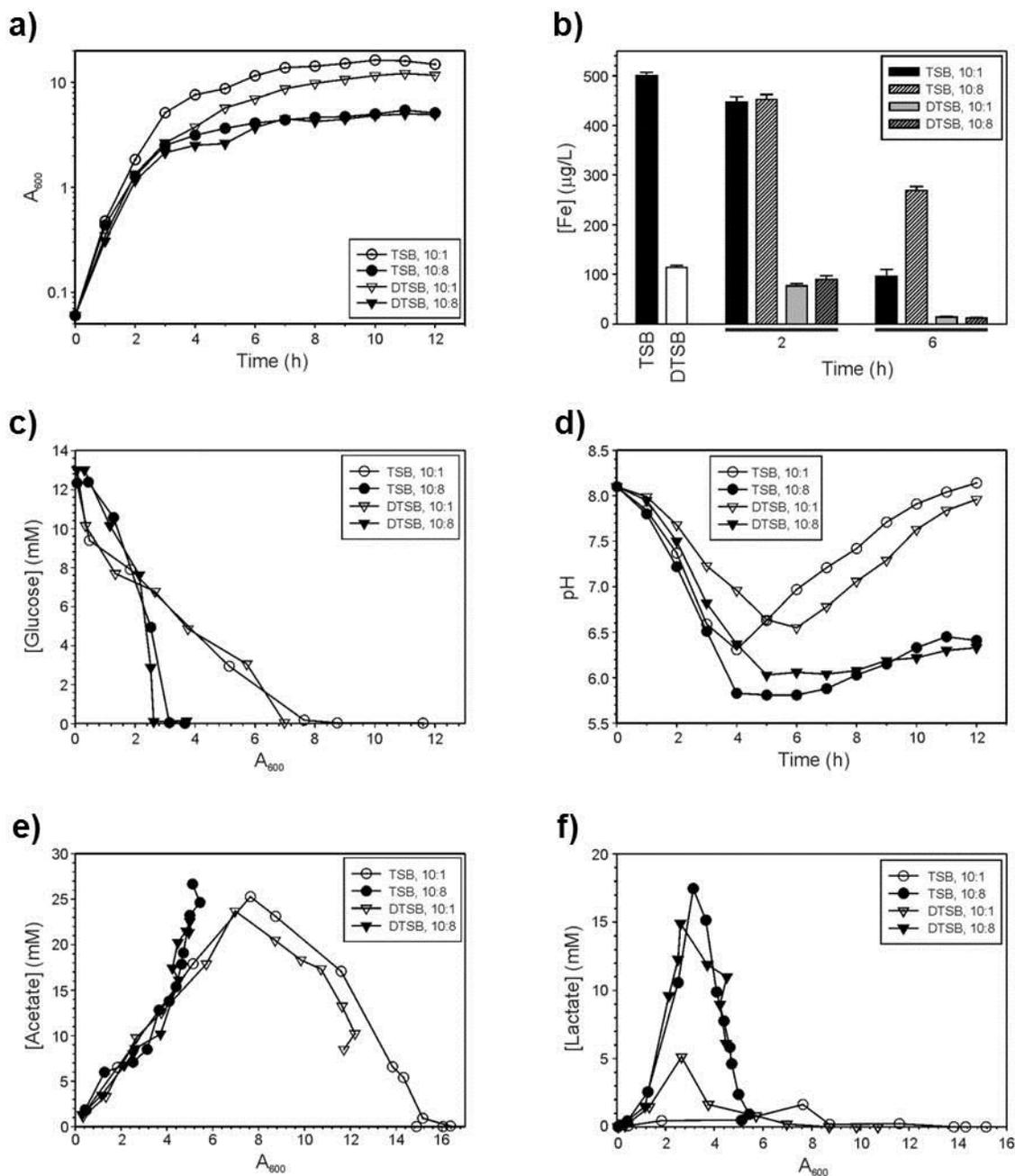


Figure 7.1 Cultivation of *S. aureus* strain SA564 under aeration- and/or iron-limitation conditions. (a) Growth curves. (b) Iron concentrations of un-inoculated media and from cultivation media harvested in the exponential and post-exponential growth phases. (c) Glucose utilization as a function of growth. (d) pH profiles. (e) Acetate accumulation and

depletion in culture media. (f) Lactate accumulation and depletion in the culture media. Data are presented as representative of experiments performed at least twice. The data in panel B are presented as the mean of 3 independent experiments done in triplicate with error bars representing standard deviations. Figure insets: Culture medium: TSB = tryptic soy broth, DTSB = deferrated TSB; Flask-to-medium ratio: 10:1 (aerobic), 10:8 (microaerobic/anaerobic).

S. aureus is a facultative anaerobe and is well-adapted for growth in low-oxygen environments when nutrients are abundant. As nutrients become scarce, *S. aureus* catabolize incompletely oxidized carbon compounds (*e.g.*, lactic acid or acetic acid) that accumulate in the medium during the exponential growth phase;²⁷ hence, the pH of the culture medium will alkalinize. As expected, strain SA564 grown under oxygen- and iron-replete conditions began to alkalinize the culture medium during the transition from the exponential growth phase to the post-exponential growth phase (Figure 7.1d). Iron-limited growth slightly delayed and decreased the alkalinization of the culture medium, while oxygen-limitation severely restricted alkalinization of the medium.

In contrast to previous speculation, iron-limitation does not lead to the acidification of the culture medium.²⁶ As stated above, the acidification of the culture medium is the result of the inefficient use of carbohydrates during growth in nutrient-rich conditions, which occurs irrespective of the availability of iron or oxygen (Figure 7.1c-1f). The decreased ability of oxygen-limited cultures, and to a lesser extent iron-limited cultures, to alkalinize the medium was most likely due to a diminished capacity to catabolize organic acids and amino acids. To address this possibility, the concentrations

of acetic acid and lactic acid in the culture media were measured throughout the growth cycle (Figure 7.1e and 1f). During microaerobic/anaerobic growth (*i.e.*, when the flask-to-medium ratio is 10:8), pyruvate is largely reduced to lactic acid (Figure 7.1f).^{49,50} In contrast, when the flask-to-medium ratio was altered to 10:1, pyruvate was oxidized to acetyl-CoA and CO₂ by the pyruvate dehydrogenase complex (Figure 7.1e).⁵¹ In the exponential growth phase, acetyl-CoA is used to generate the small phospho-donor acetyl-phosphate, which serves as a substrate for acetate kinase in substrate-level phosphorylation to generate ATP and acetic acid (Figure 7.1e). During the post-exponential growth phase, this acetic acid can be used in the synthesis of acetyl-CoA to fuel the TCA cycle. When iron was limiting, the catabolism of acetate and the biomass generated from that catabolism were reduced (Figure 7.1e). As expected, oxygen-limitation inhibited the catabolism of acetate; however, it had only a slight effect on lactate catabolism (Figure 7.1e and 1f). The latter data were surprising because acetate catabolism through the TCA cycle²⁷ and lactate catabolism through L-lactate-quinone oxidoreductase (Lqo)⁵² both require electron acceptors to balance the redox status.

7.3.2 Iron- and oxygen-limitation create metabolic blocks

Staphylococci lack a glyoxylate shunt and catabolize acetate through the tricarboxylic acid (TCA) cycle;^{27,53} hence, a diminished ability to catabolize acetate (Figure 7.1e) suggests the TCA cycle had reduced activity under iron- and oxygen-limited growth. To test this suggestion, the TCA cycle enzyme aconitase was assayed for activity under iron- and oxygen-limited growth conditions (Figure 7.2a & 2b). Since

aconitase is an iron-sulfur cluster containing enzyme, iron-limitation would be expected to decrease its activity. As expected, growth of *S. aureus* strain SA564 in iron-limited culture medium dramatically decreased post-exponential growth phase aconitase activity (Figure 7.2a). As stated earlier, one function of iron is to facilitate electron transfer reactions, a process that requires an electron donor (*e.g.*, NADH, NADPH) and acceptor (*e.g.*, oxygen). In the presence of iron but the absence of an electron acceptor, reduced dinucleotides will accumulate, which should inhibit enzyme activity through a feedback mechanism. To test this hypothesis, we limited oxygen availability by varying the flask-to-medium ratio and assayed for aconitase activity (Figure 7.2b).²⁹ As the medium volume was increased the surface area available for oxygen diffusion decreased, resulting in a post-exponential growth phase decrease in aconitase activity. While aconitase is an excellent metabolic sentinel, iron- and oxygen-limitation will affect the activity of many enzymes in bacteria.

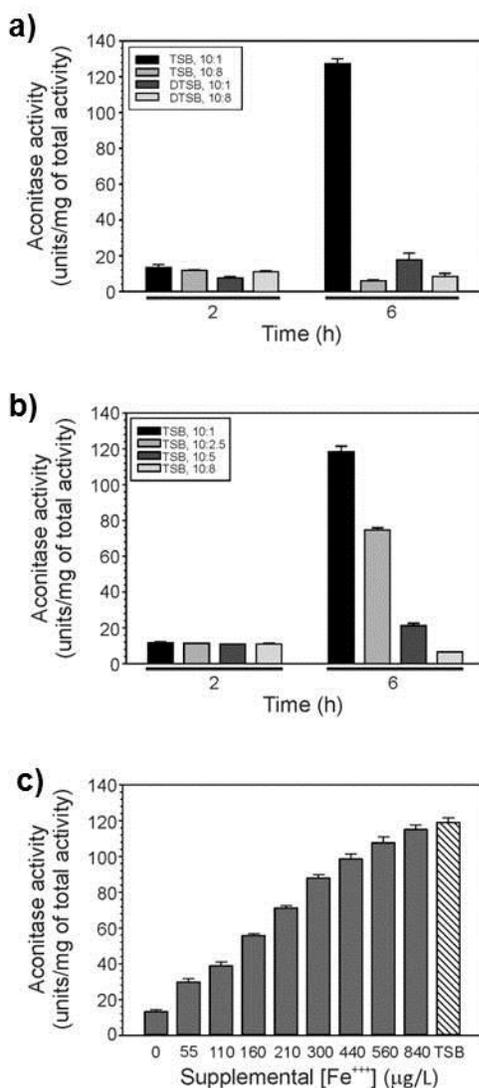


Figure 7.2 Temporal and stress related aconitase activity differences in *S. aureus* strain SA564 in the exponential (2h) and post-exponential (6h) growth phases. (a) Aconitase activity during iron-limited growth. (b) Aconitase activity with decreasing aeration flask-to-medium ratios of 10:8, 10:5, 10:2.5, 10:1. (c) Addition of iron to DTSB medium restores aconitase activity.

7.3.3 Iron and/or oxygen limitation alters the *S. aureus* metabolome

Iron- and oxygen-limited growth created a metabolic block at aconitase in the TCA cycle (Figure 7.2a & 2b); however, these stresses will also inhibit the activity of many different metabolic enzymes and pathways. It is impractical to assay for all enzymes that might be affected by iron- and/or oxygen-limited growth, so metabolic changes were assessed using NMR metabolomics (Figure 7.3). As expected, principal component analysis (PCA) of exponential growth phase metabolomes revealed they were minimally affected by iron-limitation; whereas, oxygen-limitation caused modest changes in the metabolomes (Figure 7.3a & 3c). The heterogeneity of the exponential growth phase metabolomes is easily demonstrated in a PCA shared and unique structure (SUS)-plot, where very few shared or inversely shared metabolites were identified (Figure 7.3e). In contrast to the exponential growth phase, iron-limitation significantly altered the post-exponential growth phase metabolome PCA scores plot (Figure 7.3b, 3d & 3f). Similarly, oxygen-limitation had a very pronounced effect on the metabolome. Interestingly, strain SA564 metabolomes from cultures grown in iron-limited medium and with reduced aeration clustered nearer to the aeration-limited cultures than they did with the iron-limited cultures in a PCA scores plot (Figure 7.3b). In other words, the effect of iron-limitation was masked by growth under conditions that limit the diffusion of oxygen into the culture medium. This can also be easily seen in the metabolic heat map where the metabolic changes in post-exponential growth phase (6 h) cultures grown under aeration-limitation are very similar, irrespective of iron-limitation (Figure 7.4). Taken together, these data demonstrate the significance of iron-limitation on *S. aureus* metabolism is largely determined by the growth phase and the availability of oxygen.

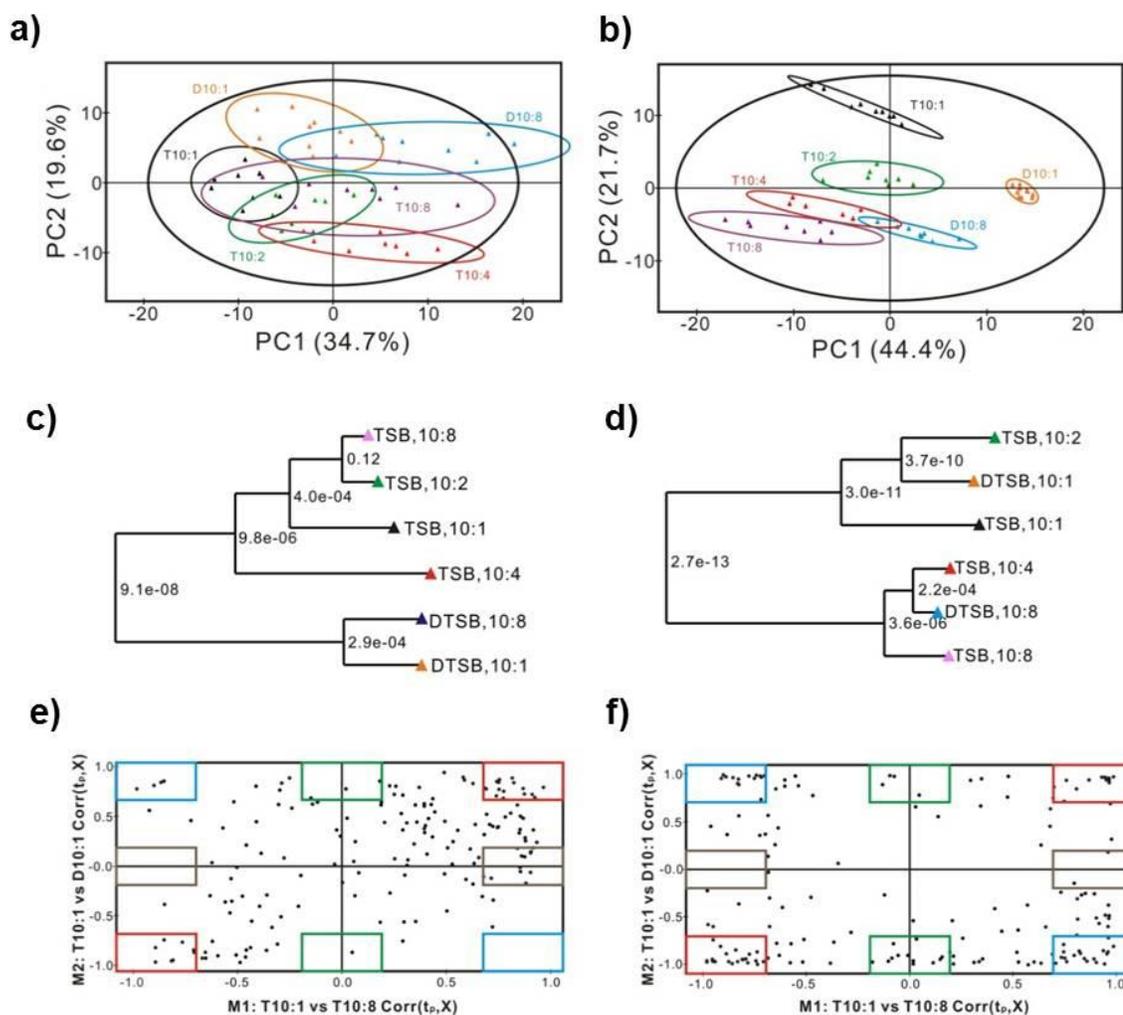


Figure 7.3 Exponential and post-exponential growth phase metabolomic changes associated with iron- and/or oxygen-limited cultivation. PCA scores plot of the (a) exponential and (b) post-exponential growth phases of iron- and/or oxygen-limited cultures of *S. aureus* strain SA564. The ellipses in the PCA scores plot correspond to the 95% confidence limits from a normal distribution for each cluster. Metabolic tree generated using the PCA scores plot data demonstrating the relationship of iron- and/or oxygen-limited cultures in the (c) exponential growth phase and (d) post-exponential growth phase. The statistical significance of each node in the metabolic tree is indicated by a p -value. Shared and unique structure (SUS)-plots generated from the PCA model of

the (e) exponential and (f) post-exponential growth phases that compares the DTSB, 10:1 and TSB, 10:8 classes to a common reference (TSB, 10:1). The red boxes highlight the chemical shift bins with intensity changes that are shared by both models. The blue boxes highlight the chemical shift bins with intensity changes that are negatively-shared by both models (relative intensity changes are in the opposite direction). The brown boxes highlight the chemical shift bins with intensity changes unique to model 1 (TSB, 10:8 vs. TSB, 10:1). The green boxes highlight the chemical shift bins with intensity changes unique to model 2 (DTSB, 10:1 vs. TSB, 10:1). In the post-exponential growth phase, 73 bins were found to be present in the shared/inversely shared region of the plot, compared to 29 bins to be found in the same regions of the exponential phase. Culture medium: TSB = tryptic soy broth, DTSB = deferrated TSB; Flask-to-medium ratio: 10:1 (aerobic), 10:8 (microaerobic/anaerobic).

To identify metabolites that contributed to the separation of the sample clusters, OPLS-DA was employed (Figure 7.4 & 1.5). In contrast to PCA, OPLS-DA is a supervised classification analysis where all the variation leading to separation between the two groups is aligned in the X-axis (P_p) and all other variation is aligned in the Y-axis (P_o).⁵⁴ To achieve separation based on aeration in an OPLS-DA scoring plot, clusters TSB, 10:1 and DTSB, 10:1 were assigned a value of 0 and the clusters of TSB, 10:8 and DTSB, 10:8 were assigned a value of 1 (Figure 7.4). This model resulted in one predictive and two orthogonal (1+2) components with a cross-validated predictive ability of $Q^2(Y)=0.883$, which indicated the data fit well within the model. The model was validated with CV-ANOVA producing a statistically significant p -value of 8.27×10^{-12} . Based on this model, the perturbations caused by iron-limitation were largely suppressed

when aeration was also limiting. As expected, the majority of the metabolic differences between aerobic growth and oxygen-limited growth include fermentation products, metabolites of pyruvate homeostasis, and amino acids (Figure 7.4). To achieve separation based on iron availability in an OPLS-DA scores plot, clusters TSB, 10:1 and TSB, 10:8 were assigned a value of 0 and the clusters of DTSB, 10:1 and DTSB, 10:8 were assigned a value of 1 (Figure 7.5). This model also produced an excellent fit that resulted in one predictive and three orthogonal (1+3) components with a cross-validated predictive ability of $Q^2(Y)=0.969$ (Figure 7.5). This model was also validated with CV-ANOVA producing a statistically significant p -value of 1.33×10^{-17} . In this model, the four metabolomic clusters are largely separated from each other, suggesting that oxygen-limitation as a stressor cannot be overwhelmed by iron-limitation. Altogether, the clustering of metabolomes in both models suggested that the effects of iron-limitation were suppressed by oxygen-limitation.

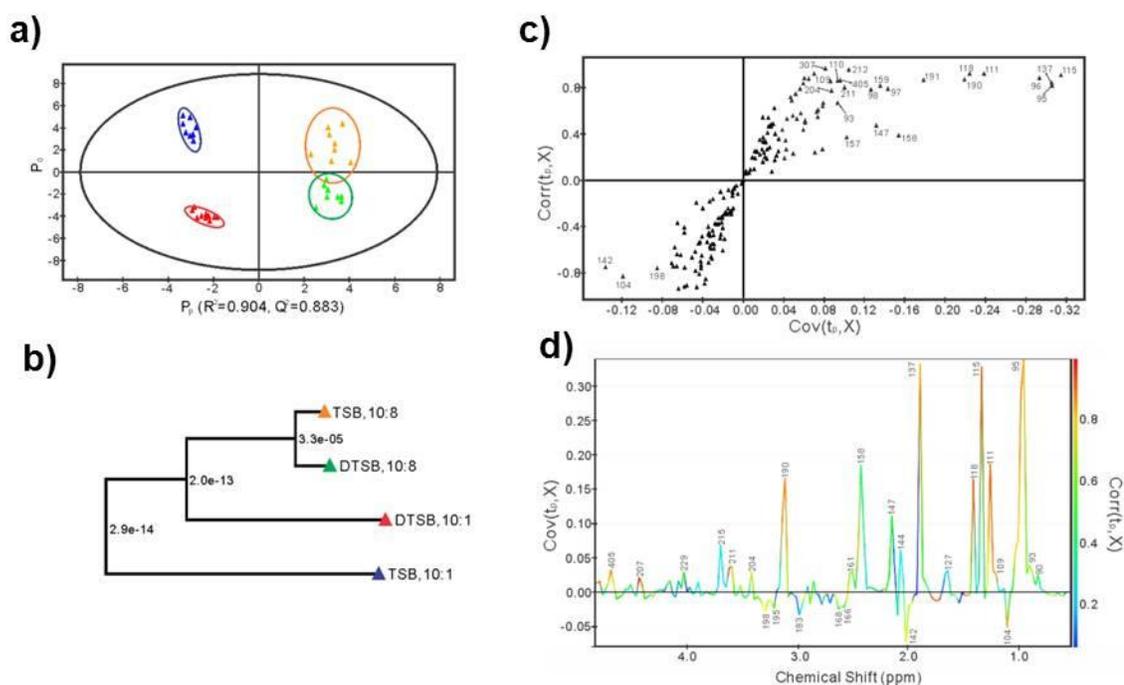


Figure 7.4 (a) OPLS-DA scores plot generated from 1D ^1H NMR spectra of iron- and/or aeration-limited cultures of *S. aureus* strain SA564 in the exponential (2 h) growth phase. The ellipses in the PCA scores plot correspond to the 95% confidence limits from a normal distribution for each cluster. TSB, 10:1 (\blacktriangle); DTSB, 10:1 (\blacktriangle); TSB, 10:8 (\blacktriangle); DTSB, 10:8 (\blacktriangle). (b) Metabolic tree generated using the OPLS-DA data demonstrating the relationship of iron- and/or oxygen-limited cultures in the post-exponential growth phase. (c) S-plot identifies the chemical shift bins that significantly contribute to class separation in the OPLS-DA scores plot. (d) OPLS-DA loading plot comparing two aeration conditions 10:1 and 10:8 for *S. aureus* strain SA564 grown on either TSB or DTSB media. Negative values indicate a decrease in peak intensity when comparing 10:8 to 10:1, while positive values indicate an increase in peak intensity. The color scale on the right indicates the relative correlation these data to the OPLS-DA model and class separation.

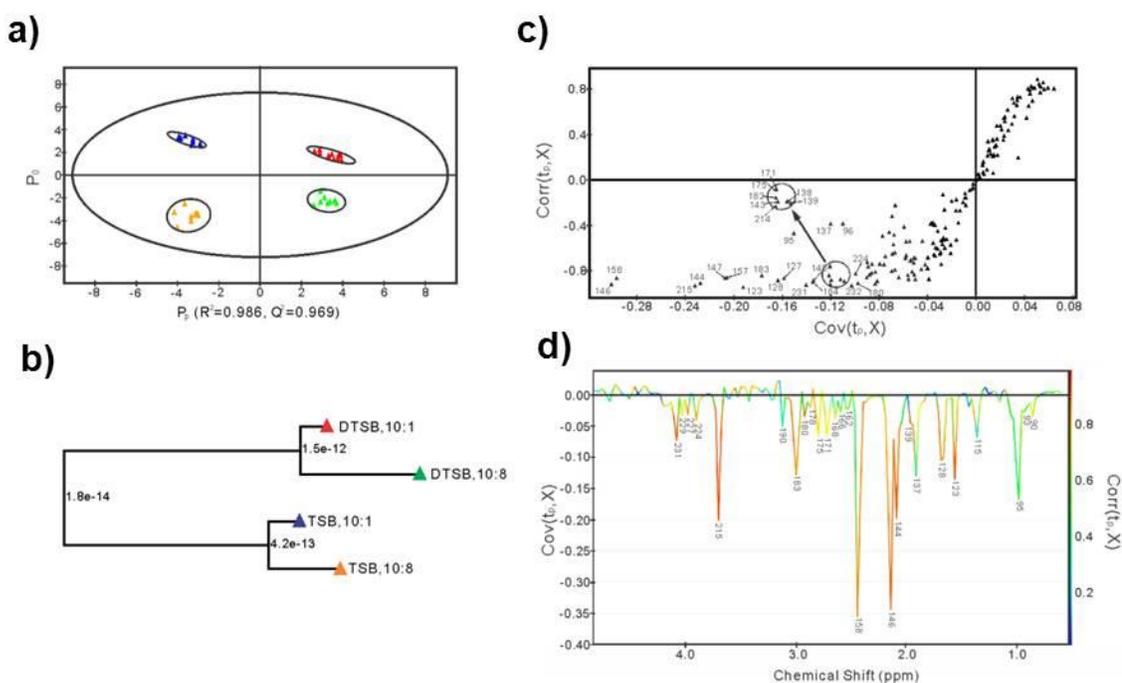


Figure 7.5 (a) OPLS-DA scores plot generated from 1D ¹H NMR spectra of iron- and/or aeration-limited cultures of *S. aureus* strain SA564 in the post-exponential (6 h) growth phase. The ellipses in the PCA scores plot correspond to the 95% confidence limits from a normal distribution for each cluster. TSB, 10:1 (▲); DTSB, 10:1 (▲); TSB, 10:8 (▲); DTSB, 10:8 (▲). (b) Metabolic tree generated using the OPLS-DA data demonstrating the relationship of iron- and/or oxygen-limited cultures in the post-exponential growth phase. (c) S-plot identifies the chemical shift bins that significantly contribute to class separation in the OPLS-DA scores plot. (d) OPLS-DA loading plot comparing two aeration conditions 10:1 and 10:8 for *S. aureus* strain SA564 grown on either TSB or DTSB media. Negative values indicate a decrease in peak intensity when comparing DTSB media to TSB media, while positive values indicate an increase in peak intensity. The color scale on the right indicates the relative correlation these data to the OPLS-DA model and class separation.

7.3.4 Iron-regulated gene transcription varies with oxygen availability

If oxygen-limitation can mask metabolic changes associated with iron-limitation, then it raises the question as to whether iron-dependent transcriptional changes can be obscured by oxygen-limitation. To address this question, bacteria were grown under iron- and/or oxygen-limited culture conditions and transcriptional changes in genes regulated by iron were assessed using qRT-PCR (Figure 7.7). As a positive control, transcription of the aconitase gene *acnA/citB* was assessed and compared to the enzyme activity data (Figure 7.2a). Overall, the *acnA/citB* mRNA levels correlated well ($r = 0.83$) with the enzymatic activity data. Similar to the metabolomics data, oxygen-limitation was able to mask the effect of iron-limited growth on some genes. Specifically, oxygen-limitation dramatically decreased the exponential (2 h) and post-exponential (6 h) growth phase transcription of the ferrous iron transporter, *feoB*, and the first gene (*i.e.*, *sbnA*) of the staphyloferrin B biosynthetic locus.⁵⁵ This reduction in transcript levels occurred despite the fact that the culture medium was iron-limited (Figure 7.1b). In contrast to *feoB* and *sbnA*, some genes [*e.g.*, *sstC*, which codes for an ATP binding cassette in a siderophore transporter]⁵⁶ had only minimal changes in mRNA levels under iron- and/or oxygen-limited growth. These data demonstrate that iron-dependent regulation of transcription in *S. aureus* can be suppressed under conditions of low oxygen-availability. In addition, these data highlight the necessity of carefully considering, and reporting, bacterial cultivation conditions.²⁹

7.4 Discussion

When cultures of *Staphylococcus epidermidis* are challenged with ethanol, antibiotics, iron-limitation, or high glucose concentrations, these stresses cause common phenotypic changes; namely, they favor biofilm formation and/or biosynthesis of polysaccharide intercellular adhesin.⁵⁷⁻⁶⁰ These same stresses cause the post-exponential growth phase metabolomic profiles to resemble that of an *S. epidermidis* TCA cycle mutant.^{17,18} The fact that divergent environmental challenges produce similar metabolic and phenotypic alterations suggests that common regulatory changes are occurring to produce the similar phenotypic outcomes.³ This suggestion was confirmed when it was observed that an *S. epidermidis* *ccpA* mutant failed to respond to metabolic changes associated with TCA cycle stress.⁷ Similar to *S. epidermidis*, we hypothesize that divergent environmental stresses cause common metabolic changes in *S. aureus* that alter the activity of metabolite-responsive regulators. The first step in identifying metabolite-responsive regulators is to define the common metabolic changes associated with environmental stresses (Figure 7.4 & 1.8). The metabolic effects of many of the environmental stresses *S. aureus* encounters have been studied in detail; however, with little, or no, deference to the temporal nature and the interconnectedness of metabolic changes.^{26,61-63} The intent of the current study was to identify common metabolic changes associated with iron- and/or oxygen-limitation that could alter the activity of metabolite-responsive regulators; the identification of these regulators is an active area of research in

our laboratories. As a side benefit, this study reinforces the necessity of carefully considering and reporting cultivation conditions.⁶⁴

While the most prominent metabolomic differences are in the post-exponential growth phase, the exponential growth phase changes are greatest under combined iron- and oxygen-limited growth (Figure 7.3, 1.4 & 7.8). As an example, both iron- and oxygen-limitation redirect carbon into amino sugar biosynthesis (*e.g.*, UDP-*N*-acetylglucosamine, glucosamine-6-phosphate, acetyl-glucosamine). This redirection of carbon appears to be at the expense of the exponential growth phase basal level TCA cycle carbon flow because the concentration of succinate is decreased relative to the iron- and oxygen-replete growth conditions (Figure 7.4 & 7.8). We observed a similar redirection of carbon into amino sugar biosynthesis in *S. epidermidis*, where TCA cycle stress induces the formation of an *N*-acetyl-glucosamine polymer known as polysaccharide intercellular adhesin.^{43,65} Together these data suggest that TCA cycle dependent regulation of polysaccharide intercellular adhesin synthesis is common to the staphylococci and not just *S. epidermidis*. While amino sugar biosynthesis was increased during oxygen- and iron-limited growth, *S. aureus* acidified the culture medium irrespective of iron or oxygen availability (Figure 7.1d). This latter observation is consistent with previous observations on staphylococcal carbohydrate catabolism,^{27,66,67} yet inconsistent with more recent speculation.²⁶

As stated above, the metabolic perturbations during iron- and/or oxygen-limited cultivation are greatest in the post-exponential growth phase (Figure 7.3a & 3b). These

differences are primarily a consequence of three factors: 1.) the TCA cycle is catabolite repressed during the exponential growth phase;²⁷ 2.) de-repression of the TCA cycle creates a large demand for iron in the post-exponential growth phase [Fig 1F and ⁵]; and 3.) TCA cycle activity generates reducing potential, which creates a need for an electron acceptor, most commonly, oxygen. Once glucose is depleted from the culture medium during growth in iron- and oxygen-replete conditions (Figure 7.1b), catabolite repression of the TCA cycle genes is relieved and enzymatic activities increase (Figure 7.2a, 2b & 1.7). In contrast, when *S. aureus* are cultivated under iron- and/or oxygen-limited conditions, transcription of TCA cycle genes and enzymatic activity remain low despite the depletion of glucose (Figure 7.1, 2a & 1.7). Though oxygen and iron are both important for post-exponential phase growth (Figure 7.1), the availability of oxygen is the major determinant of the metabolome (Figure 7.3b & 1.4). The absence of oxygen prevents carbon flow through the TCA cycle, decreasing the concentrations of biosynthetic intermediates and the amino acids derived from those intermediates (Figure 7.2a, 2b & 7.3); specifically, citric acid, succinate, and α -ketoglutarate, glutamate, glutamine, aspartate, and asparagine are decreased (Figure 7.4 & 7.8). In other words, the decreased carbon flow through the TCA cycle results in multiple amino acid auxotrophies. In addition, decreased carbon flow through the TCA cycle likely decreased the availability of oxaloacetate/PEP for gluconeogenesis, which is essential for growth when glucose has been depleted from the medium. When combined, the amino acid auxotrophies, a lack of gluconeogenesis, and inhibition of the electron transport chain

(*i.e.*, due to the lack of an electron acceptor) result in a decreased growth yield (Figure 7.1a).

Inhibition of electron transport forces bacteria to use alternative pathways to oxidize dinucleotides; however, these alternatives are strongly dependent upon carbon availability. When carbon sources that can generate pyruvate (*e.g.*, glucose and serine) are abundant in the medium, bacteria can maintain redox balance by using enzymes like lactate dehydrogenase, which leads to the accumulation of lactic acid in the medium (Figure 7.1f & 7.4). In addition to lactate dehydrogenase, *S. aureus* can oxidize dinucleotides through alanine dehydrogenase, a result consistent with the post-exponential growth phase accumulation of D-alanyl-D-alanine and alanine. These data suggest that as readily catabolizable carbon sources were being depleted from the medium several pathways are used to oxidize dinucleotides, resulting in an accumulation of NAD^+ (Figure 7.4). While oxidation of dinucleotides is important, it is only one part of redox homeostasis, the other being reduction.

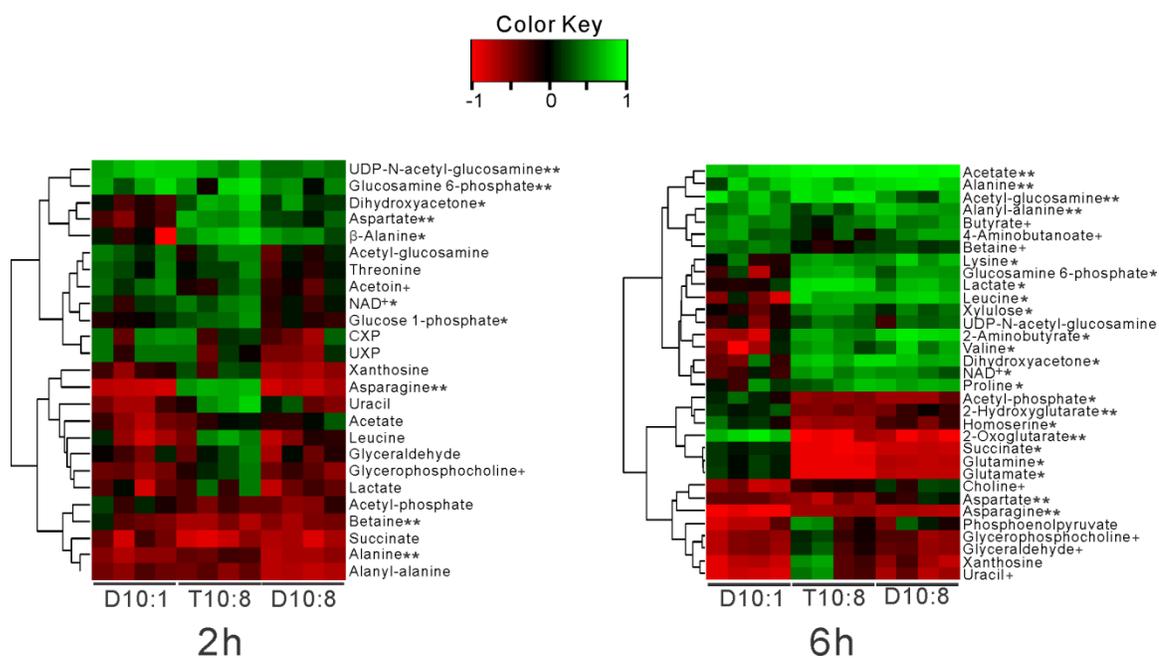


Figure 7.6 Heat maps showing the changes in metabolites as a function of cultivation conditions and growth phase. Culture medium: TSB = tryptic soy broth, DTSB = deferrated TSB; Flask-to-medium ratio: 10:1 (aerobic), 10:8 (microaerobic/anaerobic). Statistical significance ($p < 0.05$): ** = both oxygen- and iron-limitation; * = only oxygen-limitation; + = only iron-limitation.

Table 7.2. Summary of metabolite changes.^a

Culture conditions	TSB, 10:8		DTSB, 10:1		DTSB, 10:8		Metabolite	Metabolic Pathway	
	Ex	PE	Ex	PE	Ex	PE		KEGG ID	KEGG ID
Growth phase	p		p		p		KEGG ID	KEGG ID	
Xanthosine		↓		↓	↓	↓	C01762	00230, 01065	
Uracil		↑					C00106	00240,00410,00770	
UDPAcGlcN	↑↑	↑↑	↑↑	↑↑	↑	↑↑↑	C00043	00520, 00540,00550	
Succinate		↑↑	↑	↑↑		↑↑↑	C00042	00020,00190,00250,00350,00360 ,00630,00650 ,02020	
Ribose				↓		↓	C00121	00030,02010	
Proline		↑↑		↑		↑↑↑	C00148	00330,00970 ,01230,02010	
NADH		↓		↓		↓	C00004	00190, 01100, 04020	
NAD ⁺	↑↑	↑↑				↑	C00003	00190	
Lysine		↑↑				↑	C00047	00300,00310,00780, 00970,01210,01230,02010	
Lactate		↑↑					C00186	00010,00620,00640	
GPC		↓	↓	↓↓		↓↓	C00670	00564,00565	
Glyceraldehyde		↑↑					C00577	00030,00051,00561	
Glutamine				↑↑		↑↑↑	C00064	00230,00240,00250,00330,00471,00970,01230, 02010,02020	
Glutamate				↑↑		↑↑↑	C00025	00250,00330,00340,00430,00471,00480,00650, 00660,00970,01210,01230,02010,02020	
Glc6P	↑	↑↑	↑			↑↑↑	C00352	00250,00520,02060	
Dihydroxyacetone	↑↑	↑↑				↑↑↑	C00184	00561,00680	
Betaine	↓↓	↓↓	↓		↓↓	↓	C00719	00260,02010	
Aspartate	↑↑		↓	↑↑		↑↑↑	C00049	00250,00260,00270,00300,00330,00340,00410, 00770,00970,01210,01230,02010,02020	
Asparagine	↑↑	↑↑	↓↓		↓↓	↑↑↑	C00152	00250,00970,01230	
Alanyl-alanine	↑	↑	↓	↑↑	↓	↑↑↑	C00993	00473,00550	
Alanine	↓		↓↓		↓↓	↑↑↑	C00041	00250,00270,00430 ,00473,00970,01230,02010	
Acetyl-phosphate		↓↓		↓	↓	↓↓	C00227	00430,00620,00680	
Acetyl-glucosamine				↓		↓	C00140	00520, 01110,02010,02060	
Acetate		↑↑		↓		↑↑↑	C00033	00010,00430,00440,00534,00620,00660,00680	
4-Aminobutyrate				↑		↑	C00334	00250,00330,00410,00650	
2-Hydroxyglutarate		↑↑		↑↑		↑↑	C00026	00020,00040,00053,00250,00300,00340,00430, 00471,00650,00660,01210,01230	

^a“↑” indicates an increase; “↓” indicates a decrease. One arrow represents a $p \leq 0.05$ (Student’s T-test), two arrows represents a $p \leq 0.01$, three 3 arrows represents a $p \leq 0.001$.

^bMetabolites altered in the exponential (Exp) growth phase in TSB, 10:8, DTSB,10:1 and DTSB10:8 relative to the cultures grown in TSB with a 10:1 flask-to-medium ratio.

^cMetabolites altered in the post-exponential phase (PE) of growth in TSB, 10:8, DTSB,10:1 and DTSB10:8 relative to the cultures grown in TSB with a 10:1 flask-to-medium ratio.

^dKEGG (Kyoto Encyclopedia of Genes and Genomes) database identification number for the metabolite and the associate metabolic pathways.

When the TCA cycle is inhibited by iron- and/or oxygen-limitation, the ability to reduce dinucleotides and restore redox balance is also greatly impaired. This inability to reduce dinucleotides is reflected in the decreased concentrations of TCA cycle associated metabolites and amino acids, namely, aspartate asparagine, citrate, succinate, α -ketoglutarate, glutamate, and glutamine. In addition to the TCA cycle, glycerol-3-phosphate can be used in an electron transfer reaction to reduce NAD^+ to NADH and generate dihydroxyacetone phosphate. This reduction reaction is consistent with the oxygen-limited accumulation of dihydroxyacetone (the phosphate is labile and can be lost during harvest and sample preparation to generate the more stable dihydroxyacetone) (Figure 7.4). When carbon flow through the TCA cycle and pentose phosphate pathway is decreased, the ability to generate reducing potential via the NAD(P)H-dependent glycerol-3-phosphate dehydrogenase (Figure 7.4) can be important for redox homeostasis and biosynthesis. That being said, this metabolic rearrangement is insufficient to offset the accumulation of NAD^+ . In summary, the necessity of maintaining redox homeostasis explains why the availability of oxygen (*i.e.*, its function as an electron acceptor) has a greater influence on the post-exponential growth phase metabolome than does the availability of iron (Figure 7.3b). In the absence of an electron acceptor, iron is not

needed for electron transport; hence, the demand for iron will be lower. This also explains why oxygen availability alters transcription of genes involved in iron-acquisition (Figure 7.7). In essence, decreased oxygen availability suppresses the iron-sparing response, a phenomenon similar to that seen in *Saccharomyces cerevisiae*.⁶⁸

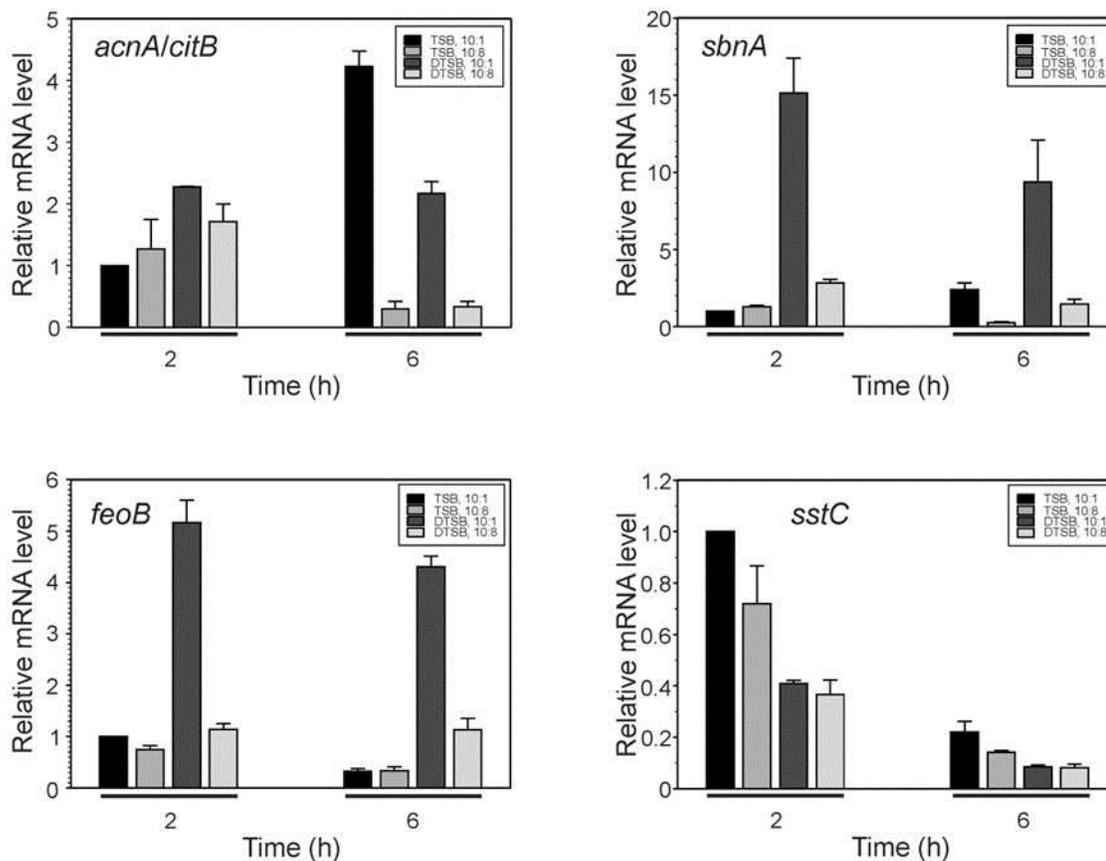


Figure 7.7 Oxygen-limitation influences transcription of iron-regulated genes in the exponential (2 h) and post-exponential (6 h) growth phases. Relative mRNA levels for *acnA/citB*, *sbnA*, *feoB*, and *sstC* as determined by qRT-PCR. The data are the mean and SEM of at least 2 biological replicates each determined in duplicate. Culture medium: TSB = tryptic soy broth, DTSB = deferrated TSB; Flask-to-medium ratio: 10:1 (aerobic), 10:8 (microaerobic/anaerobic).

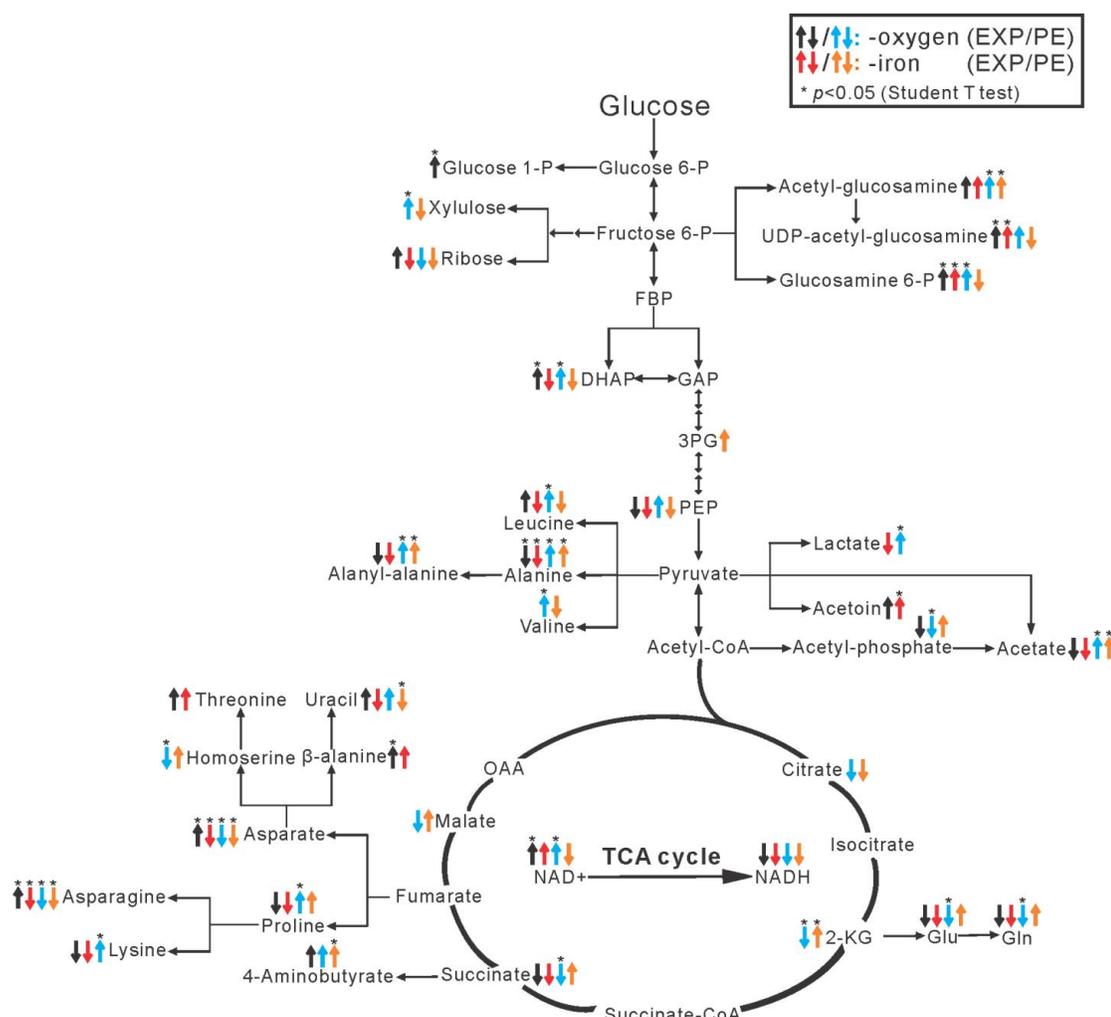


Figure 7.8 Schematic representation of the central metabolic changes associated with iron- or oxygen-limitation. The black/blue arrows indicate exponential/post-exponential growth phase increased or decreased intracellular concentrations of metabolites in strain SA564 cultivated in TSB with a 10:8 flask-to-medium ratio relative to a control culture grown in TSB with a flask-to-medium ratio of 10:1. The red/orange arrows indicate exponential/post-exponential growth phase increased or decreased intracellular concentrations of metabolites in strain SA564 grown in DTSB with a flask-to-medium ratio of 10:1 relative to a control culture of *S. aureus* strain SA564 grown in TSB with a flask-to-medium ratio of 10:1. Statistical significance at the 95% confidence level ($p < 0.05$) is denoted by asterisks above the arrows.

7.5 Reference

- (1) Vandenesch, F.; Kornblum, J.; Novick, R. P. *J Bacteriol* **1991**, *173*, 6313.
- (2) Richardson, A. R.; Libby, S. J.; Fang, F. C. *Science* **2008**, *319*, 1672.
- (3) Somerville, G. A.; Proctor, R. A. *Microbiol Mol Biol Rev* **2009**, *73*, 233.
- (4) Somerville, G. A.; Chaussee, M. S.; Morgan, C. I.; Fitzgerald, J. R.; Dorward, D. W.; Reitzer, L. J.; Musser, J. M. *Infect Immun* **2002**, *70*, 6373.
- (5) Somerville, G. A.; Cockayne, A.; Dür, M.; Peschel, A.; Otto, M.; Musser, J. M. *J Bacteriol* **2003**, *185*, 6686.
- (6) Majerczyk, C. D.; Dunman, P. M.; Luong, T. T.; Lee, C. Y.; Sadykov, M. R.; Somerville, G. A.; Bodi, K.; Sonenshein, A. L. *J Bacteriol* **2010**, *192*, 2861.
- (7) Sadykov, M. R.; Hartmann, T.; Mattes, T. A.; Hiatt, M.; Jann, N. J.; Zhu, Y.; Ledala, N.; Landmann, R.; Herrmann, M.; Rohde, H.; Bischoff, M.; Somerville, G. A. *Microbiology* **2011**, *157*, 3458.
- (8) Zhu, Y.; Nandakumar, R.; Sadykov, M. R.; Madayiputhiya, N.; Luong, T. T.; Gaupp, R.; Lee, C. Y.; Somerville, G. A. *J Bacteriol* **2011**, *193*, 6187.
- (9) Majerczyk, C. D.; Sadykov, M. R.; Luong, T. T.; Lee, C.; Somerville, G. A.; Sonenshein, A. L. *J Bacteriol* **2008**, *190*, 2257.
- (10) Pohl, K.; Francois, P.; Stenz, L.; Schlink, F.; Geiger, T.; Herbert, S.; Goerke, C.; Schrenzel, J.; Wolz, C. *J Bacteriol* **2009**.
- (11) Deutscher, J.; Kuster, E.; Bergstedt, U.; Charrier, V.; Hillen, W. *Mol Microbiol* **1995**, *15*, 1049.

- (12) Seidl, K.; Stucki, M.; Ruegg, M.; Goerke, C.; Wolz, C.; Harris, L.; Berger-Bachi, B.; Bischoff, M. *Antimicrob Agents Chemother* **2006**, *50*, 1183.
- (13) Jaeger, T.; Mayer, C. *J Bacteriol* **2008**, *190*, 6598.
- (14) Shivers, R. P.; Sonenshein, A. L. *Mol Microbiol* **2004**, *53*, 599.
- (15) Romling, U.; Galperin, M. Y.; Gomelsky, M. *Microbiol Mol Biol Rev* **2013**, *77*, 1.
- (16) Ernst, J. F.; Bennett, R. L.; Rothfield, L. I. *J Bacteriol* **1978**, *135*, 928.
- (17) Sadykov, M. R.; Zhang, B.; Halouska, S.; Nelson, J. L.; Kreimer, L. W.; Zhu, Y.; Powers, R.; Somerville, G. A. *J. Biol. Chem.* **2010**, *285*, 36616.
- (18) Zhang, B.; Halouska, S.; Schiaffo, C. E.; Sadykov, M. R.; Somerville, G. A.; Powers, R. *J Proteome Res* **2011**, *10*, 3743.
- (19) Cohen, S.; Sweeney, H. M.; Leitner, F. *J Bacteriol* **1967**, *93*, 1227.
- (20) Johnson, M.; Cockayne, A.; Morrissey, J. A. *Infect Immun* **2008**, *76*, 1756.
- (21) Trivier, D.; Courcol, R. J. *FEMS Microbiol Lett* **1996**, *141*, 117.
- (22) Dassy, B.; Fournier, J. M. *Infect Immun* **1996**, *64*, 2408.
- (23) Jacobs, N. J.; Maclosky, E. R.; Conti, S. F. *J Bacteriol* **1967**, *93*, 278.
- (24) Ferreira, M. T.; Manso, A. S.; Gaspar, P.; Pinho, M. G.; Neves, A. R. *PLoS One* **2013**, *8*, e58277.
- (25) Fuchs, S.; Pane-Farre, J.; Kohler, C.; Hecker, M.; Engelmann, S. *J Bacteriol* **2007**, *189*, 4275.

- (26) Friedman, D. B.; Stauff, D. L.; Pishchany, G.; Whitwell, C. W.; Torres, V. J.; Skaar, E. P. *PLoS Pathog* **2006**, *2*, e87.
- (27) Somerville, G. A.; Sa il-Salim, B.; Wickman, J. M.; Raffel, S. J.; Kreiswirth, B. N.; Musser, J. M. *Infect Immun* **2003**, *71*, 4724.
- (28) Sheppard, L. N.; Kontoghiorghes, G. J. *Arzneimittel-Forschung* **1993**, *43*, 659.
- (29) Somerville, G. A.; Proctor, R. A. *BMC Microbiol* **2013**, *13*, 9.
- (30) Zhang, B.; Halouska, S.; Gaupp, R.; Lei, S.; Snell, E.; Fenton, R. J.; Barletta, R. G.; Somerville, G. A.; Powers, R. *Journal of Integrated OMICS* **2013**, *3*.
- (31) Hu, K.; Westler, W. M.; Markley, J. L. *J Am Chem Soc* **2011**, *133*, 1662.
- (32) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J Biomol NMR* **1995**, *6*, 277.
- (33) *Using NMRView to Visualize and Analyze the NMR Spectra of Macromolecules*; A.K.Downling, Ed.; Humana Press, Inc., NJ: Totowa, 2004; Vol. 278.
- (34) Shao, J. *J Am Stat Assoc* **1993**, *88*, 486.
- (35) Golbraikh, A.; Tropsha, A. *Journal of molecular graphics & modelling* **2002**, *20*, 269.
- (36) Eriksson, L.; Trygg, J.; Wold, S. *J. Chemometrics* **2008**, *22*, 594.
- (37) Werth, M. T.; Halouska, S.; Shortridge, M. D.; Zhang, B.; Powers, R. *Anal Biochem* **2009**.
- (38) Worley, B.; Halouska, S.; Powers, R. *Anal Biochem* **2013**, *433*, 102.

- (39) Wishart, D. S.; Knox, C.; Guo, A. C.; Eisner, R.; Young, N.; Gautam, B.; Hau, D. D.; Psychogios, N.; Dong, E.; Bouatra, S.; Mandal, R.; Sinelnikov, I.; Xia, J.; Jia, L.; Cruz, J. A.; Lim, E.; Sobsey, C. A.; Shrivastava, S.; Huang, P.; Liu, P.; Fang, L.; Peng, J.; Fradette, R.; Cheng, D.; Tzur, D.; Clements, M.; Lewis, A.; De Souza, A.; Zuniga, A.; Dawe, M.; Xiong, Y.; Clive, D.; Greiner, R.; Nazyrova, A.; Shaykhtudinov, R.; Li, L.; Vogel, H. J.; Forsythe, I. *Nucleic Acids Res* **2009**, *37*, D603.
- (40) Akiyama, K.; Chikayama, E.; Yuasa, H.; Shimada, Y.; Tohge, T.; Shinozaki, K.; Hirai, M. Y.; Sakurai, T.; Kikuchi, J.; Saito, K. *In Silico Biology* **2008**, *8*, 339.
- (41) Markley, J. L.; Ulrich, E. L.; Berman, H. M.; Henrick, K.; Nakamura, H.; Akutsu, H. *J Biomol NMR* **2008**, *40*, 153.
- (42) Team, R. D. C. *R: A language and environment for statistical computing*; R Foundation for Statistical Computing: Vienna, 2011.
- (43) Sadykov, M. R.; Olson, M. E.; Halouska, S.; Zhu, Y.; Fey, P. D.; Powers, R.; Somerville, G. A. *J Bacteriol* **2008**, *190*, 7621.
- (44) Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguchi, A.; Aoki, K.; Nagai, Y.; Lian, J.; Ito, T.; Kanamori, M.; Matsumaru, H.; Maruyama, A.; Murakami, H.; Hosoyama, A.; Mizutani-Ui, Y.; Takahashi, N. K.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; Hirakawa, H.; Kuhara, S.; Goto, S.; Yabuzaki, J.; Kanehisa, M.; Yamashita, A.; Oshima, K.; Furuya, K.; Yoshino, C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K. *Lancet* **2001**, *357*, 1225.
- (45) Fox, J. B.; Holtman, D. F. *J Bacteriol* **1968**, *95*, 1548.
- (46) Theodore, T. S.; Schade, A. L. *J. Gen. Microbiol.* **1965**, *40*, 385.
- (47) Theodore, T. S.; Schade, A. L. *J Gen Microbiol* **1965**, *39*, 75.
- (48) Pasteur, L. *Compt. Rend. Acad. Sci. (Paris)* **1861**, *52*, 344.

- (49) Kendall, A. I.; Friedemann, T. E.; Ishikawa, M. *J. Infect. Dis.* **1930**, *47*, 223.
- (50) Krebs, H. A. *Biochem. J.* **1937**, *31*, 661.
- (51) Gardner, J. F.; Lascelles, J. *J. Gen. Microbiol.* **1962**, *29*, 157.
- (52) Fuller, J. R.; Vitko, N. P.; Perkowski, E. F.; Scott, E.; Khatri, D.; Spontak, J. S.; Thurlow, L. R.; Richardson, A. R. *Frontiers in cellular and infection microbiology* **2011**, *1*, 19.
- (53) Goldschmidt, M. C.; Powelson, D. M. *Arch. Biochem. Biophys.* **1953**, *46*, 154.
- (54) Bylesjo, M.; Rantalainen, M.; Cloarec, O.; Nicholson, J. K.; Holmes, E.; Trygg, J. *Journal of Chemometrics* **2006**, *20*, 341.
- (55) Beasley, F. C.; Cheung, J.; Heinrichs, D. E. *BMC Microbiol* **2011**, *11*, 199.
- (56) Morrissey, J. A.; Cockayne, A.; Hill, P. J.; Williams, P. *Infect Immun* **2000**, *68*, 6281.
- (57) Knobloch, J. K.; Bartscht, K.; Sabottke, A.; Rohde, H.; Feucht, H. H.; Mack, D. *J Bacteriol* **2001**, *183*, 2624.
- (58) Mack, D.; Siemssen, N.; Laufs, R. *Infect Immun* **1992**, *60*, 2048.
- (59) Deighton, M.; Borland, R. *Infect Immun* **1993**, *61*, 4473.
- (60) Rachid, S.; Ohlsen, K.; Witte, W.; Hacker, J.; Ziebuhr, W. *Antimicrob Agents Chemother* **2000**, *44*, 3357.
- (61) Nychas, G. J.; Tranter, H. S.; Brehm, R. D.; Board, R. G. *J Appl Bacteriol* **1991**, *70*, 344.

- (62) Seidl, K.; Muller, S.; Francois, P.; Kriebitzsch, C.; Schrenzel, J.; Engelmann, S.; Bischoff, M.; Berger-Bachi, B. *BMC Microbiol* **2009**, *9*, 95.
- (63) Ulrich, M.; Bastian, M.; Cramton, S. E.; Ziegler, K.; Pragman, A. A.; Bragonzi, A.; Memmi, G.; Wolz, C.; Schlievert, P. M.; Cheung, A.; Doring, G. *Mol Microbiol* **2007**, *65*, 1276.
- (64) Neidhardt, F. C. *Nat Rev Microbiol* **2006**, *4*, 876.
- (65) Vuong, C.; Kidder, J. B.; Jacobson, E. R.; Otto, M.; Proctor, R. A.; Somerville, G. A. *J Bacteriol* **2005**, *187*, 2967.
- (66) Blumenthal, H. J. In *The Staphylococci*; Cohen, J. O., Ed.; Wiley-Interscience: New York, 1972, p 111.
- (67) Elek, S. D. In *Staphylococcus pyogenes and its relation to disease*; E. & S. Livingstone LTD.: Edinburgh and London, 1959, p 54.
- (68) Kaplan, J.; McVey Ward, D.; Crisp, R. J.; Philpott, C. C. *Biochim Biophys Acta* **2006**, *1763*, 646.

CHAPTER 8

CCPE IS A LYSR-TYPE REGULATOR CONTROLLING TCA CYCLE ACTIVITY IN *STAPHYLOCOCCUS AUREUS*

8.1 Introduction

Carbon catabolite repression (CCR) in bacteria is a widespread, regulatory phenomenon that represses transcription of genes and operons involved in the catabolism of non-preferred carbon sources when the preferred carbon source(s) are present. CCR has been studied extensively in *Bacillus subtilis* and serves as the prototype of CCR-regulated gene expression in Gram-positive bacteria (reviewed in references 1).^{1,2} In *B. subtilis*, the catabolite control protein A (CcpA) acts in concert with the small phosphocarrier proteins histidine-containing protein (HPr) and catabolite repression HPr (Crh) to regulate transcription in response to carbohydrate availability. In addition, the metabolite-activated bifunctional HPr kinase/phosphorylase is involved in CCR through its action of phosphorylating and dephosphorylating Hpr.^{3,4} There are several other proteins in *B. subtilis* that contribute to CCR either in cooperation with, or independently of CcpA, including CcpC,^{5,6} CcpN,⁷ CitR,⁸ Crh,⁹ CodY,¹⁰ and GlcU¹¹ (reviewed in reference 1a).¹ In *Staphylococcus aureus*, homologs of CcpA and CodY regulate the transcription of numerous metabolic and biosynthetic genes and virulence determinants; thereby, linking staphylococcal carbon metabolism with pathogenicity.¹²

To identify additional CCR elements in *S. aureus*, we compared the genomes of *S. aureus* strain Newman with that of *B. subtilis* strain 168 and found that *S. aureus* had uncharacterized homologs of *B. subtilis* genes that are known to affect the regulation of

carbon catabolism. Here we report about the identification of a putative carbon catabolite responsive regulator, CcpE (NWMN_0641) that affects central metabolism by regulating tricarboxylic acid (TCA) cycle activity via transcriptional control of the aconitase-encoding gene *citB*.

8.2 Methods and materials

8.2.1 Bacterial strains and culture conditions

The bacterial strains and plasmids used in this study are listed in Table 8.1. *S. aureus* strains were grown in Luria-Bertani Lennox (LB-L) medium (BD, Heidelberg, Germany) and *B. subtilis* strains were grown in TSS minimal medium supplemented with 0.2 % (w/v) glutamine and 0.5 % (w/v) glucose (6). All strains were grown at 37 °C and aerated at 230 rpm with a flask-to-medium volume ratio of 10:1. Antibiotics, when used, were added to the medium at the following concentrations (per milliliter): 10 µg chloramphenicol, 50 µg kanamycin, and 8 µg tetracycline.

Table 8.1 Strains and plasmids used in this study.

Strain or plasmid	Relevant genotype or characteristic(s) ^a	Reference or source
<i>S. aureus</i>		
RN4220	NCTC 8325-4 r ⁻ m ⁺ (restriction-negative, modification-positive)	(34) ¹³
Newman	Clinical isolate (ATCC 25904); CP-5 producer	(35) ¹⁴
THa	RN4220 Δ <i>ccpE::lox66-aphAIII-lox71</i> , Kan ^r	This study
TH01	Newman Δ <i>ccpE::lox72</i>	This study
TH01c	TH01 harboring plasmid pTH2c <i>cis</i> -integrated at the NWMN_0640 locus, leading to a duplication of the NWMN_0640 gene, <i>ccpE</i> ^r , Tc ^r	This study
<i>B. subtilis</i>		
AF21	Δ <i>amyE::Φ(citBp21-lacZ cat)</i> , Cm ^r	(43) ¹⁵
CJB9	Δ <i>amyE::Φ(citBp21-lacZ cat) ccpC::spe</i> , Cm ^r , Spc ^r	(3b) ⁶
<i>E. coli</i>		
DH5	<i>hsdR17</i> (r ⁻ m ⁺) <i>deoR endA1 gyrA96 recA1 relA1 supE44 thi-1 F⁻ λ⁻ Δ(lacZYA-argF) U169 φ80 lacZΔM15</i>	Invitrogen
SCS110	<i>dam dcm rpsL</i> (Str ^r) <i>thr leu endA thi-1 lacY galK galT ara tonA tsx supE44 (lac-proAB) [F⁺ traD36 proAB lacI^rZ M15]</i>	Stratagene
Plasmids		
pBT	1.6-kb PCR fragment of the <i>tet</i> (L) gene of pHY300PLK into Alw261-digested pBC SK(+) (Stratagene); Tc ^r	(37) ¹⁶
pBT2-arcA	Allelic replacement vector for <i>S. aureus arcA</i> , harboring the <i>lox66-aphAIII-lox71</i> cassette; Kan ^r	(10) ¹⁷
pBT <i>lox-aph</i>	pBT with a 1.6 kb <i>lox66-aphAIII-lox71</i> fragment obtained from pBT2-arcA cloned into the PstI site of pBT; Kan ^r , Tc ^r	This study
pBus1	<i>E. coli</i> - <i>S. aureus</i> shuttle plasmid with multicloning site from pBluescript II SK (Stratagene) and the <i>rrnT14</i> terminator sequence from pLL2443; Tc ^r	(11) ¹⁸
pRAB1	<i>cat, bla, PpagA-cre</i> ; pBT2 derivative; expression of Cre in staphylococci; Cm ^r	(10) ¹⁷
pT7HMT	pET28-based expression vector, allowing a T7 polymerase-driven expression of hexahistidine-tagged fusion proteins, Kan ^r	(38) ¹⁹
pT7HMT- <i>ccpE</i>	pT7HMT derivative harboring the <i>ccpE</i> ORF cloned into the Sall and NotI sites of the vector, Kan ^r	This study
pTH2	pBT with a 3 kb fragment covering the NWMN_0641 flanking regions and the <i>lox66-aphAIII-lox71</i> resistance cassette fully replacing the <i>ccpE</i> ORF, Tc ^r	This study
pTH2c	pBT with a 2 kb fragment covering the <i>ccpE</i> ORF and 1 kb of the upstream region including the NWMN_0640 ORF, Tc ^r	This study
pTH3	pBus1 with a 1.7-kb fragment covering the <i>B. subtilis ccpC</i> promoters P1/P2 and the <i>ccpC</i> ORF, Tc ^r	This study
pTH4	pBus1 with a 1.7-kb fragment covering the <i>B. subtilis ccpC</i> promoters P1/P2 and the <i>ccpE</i> ORF, Tc ^r	This study

^a Abbreviations: Cm^r, chloramphenicol resistant; Kan^r, kanamycin resistant; Spc^r, spectinomycin resistant; Tc^r, tetracycline resistant; ORF, open reading frame

Table 8.2 Primers used in this study.

Primer	Sequence (5'-3')	
Construction primer		
MBH152	forward	GCAAATTGAGCTCTATCTTTAGAGC
MBH153	reverse	gtcgGATCCGAGCATGTTGCAATTGCC
MBH154	forward	gtcggtagCTGGCGTCGCCTAATTGATAGG
MBH155	reverse	gtcctcgaGTATGGCTGCAGCTTGAATTAC
MBH225	forward	gtcggtagCAGAATAGATTTGATGCTTCAGC
MBH226	reverse	gtcggtagACCAACTACGTTCAATTTAACCG
MBH239	forward	ggggtcgacACCAACTACGTTCAATTTAACCG
MBH255	reverse	gtcggtagCGATTTTCAATATCATATGTATCAC
MBH338	forward	GCATGTFCACCTCCTTTTGTATCAATAAG
MBH341	reverse	ctgtctAGATCAACGATATAGAGACAGGG
5'-RACE primer		
5' RACE-Adapter		GAUAUGC GCGAAUCCUGUAGAACGAACACUAGAAGAAA
NWMN_0640-RT		TCTACTGTACACGCATTACC
citB-RT		ACGTCATCCATTGCTTTACG
RACE PCR 5'		GATATGCGCGAATTCCTGTAGAACG
NWMN_0640-PCR		CGTGACTAAGCCTATAATACCC
Cit-PCR		GTAATACCTTGCTCTTCTACAGC
Northern probe primer		
MBH279	forward	GGTATTATAGGCTTAGTCACG
MBH280	reverse	GCAGTACATGATTTTCTTTGG
MBH281	forward	AGACGAAACGAAAACGTTACG
MBH282	reverse	GTAACCTTTGTAAACATCATCTCG
Real-time RT PCR primer		
<i>Bs ccpC</i>	forward	TGTGAAAAAGTATCCGAATGCA
<i>Bs ccpC</i>	reverse	CTGCTCCACCCGGTTATGA
<i>ccpE</i>	forward	CCGCTGATTCGTTTCGACAT
<i>ccpE</i>	reverse	CAATTGCAACATGCTCGGAT
<i>citB</i>	forward	CAAGATCATCAAGTGCCATTTCGT
<i>citB</i>	reverse	CGTGATTACCACGTTTGAACC
<i>Bs citB</i>	forward	TAATCGGCGCGCGAAACTTGTC
<i>Bs citB</i>	reverse	CAGTAGCTATTGTTCCGTTTGG
<i>citZ</i>	forward	CCGTAGGTTCTCTGAAAGGGC
<i>citZ</i>	reverse	AACATCGTCATAACTTGTTCGTTTG
<i>gyrB</i>	forward	GACTGATGCCGATGTGGA
<i>gyrB</i>	reverse	AACGGTGGCTGTGCAATA
<i>Bs gyrB</i>	forward	CATTGGCGAAACGGATCATA
<i>Bs gyrB</i>	reverse	GGGTCCGGGACAAAATGTG
EMSA primer		
<i>ccpE</i>	forward	CAAATAAAGCCTAAGCAATAT
<i>ccpE</i>	reverse	AGTAAACGATAGTCTTCAATC
<i>citBp</i>	forward	ATCATTCTGTCCCACTCCCATC
<i>citBp</i>	reverse	GTAAGTATAACTTTGGCCATTCAAGTC
<i>citZp</i>	forward	ATTGTAATAACTTTCATGGATTATCAC
<i>citZp</i>	reverse	CTAAACCTCTTTGTAATTCTGCCAT
NWMN-0640P	forward	GTGATCAATGAAGTGGTTTGAAGG
NWMN-0640P	reverse	CTTTTACCTCTGTGTATTCCTATC

8.2.2 Construction of a *S. aureus* Δ *ccpE* mutant

1 kb fragments, containing the flanking regions of the *ccpE* (NWMN_0641) gene were amplified by PCR from chromosomal DNA of *S. aureus* strain Newman using primer-pairs MBH154/MBH155 and MBH152/MBH153, respectively (primer sequences are listed in Table 8.1). The PCR products were digested with *KpnI/XhoI* and *BamHI/SacI*, respectively, and cloned together with the *XhoI/BamHI* digested *lox66-aphAIII-lox71* resistance cassette (obtained from pBT2-arcA) into suicide vector pBT to generate plasmid pTH2. Plasmid pTH2 was electroporated into *S. aureus* strain RN4220 to obtain strain THa, in which the *ccpE* gene was replaced by the *lox66-aphAIII-lox71* cassette by allelic replacement. THa was then used as a donor for transducing the *ccpE* deletion into other *S. aureus* strains. The marker-free Δ *ccpE* mutant TH01 of *S. aureus* strain Newman was constructed by treatment with the Cre recombinase expressed from the temperature-sensitive vector pRAB1,¹⁷ which was subsequently removed from TH01 by culturing the strain at 42 °C. The deletion of *ccpE* in TH01 was confirmed by PCR, pulsed-field gel electrophoresis (PFGE) of total genome *SmaI* digests, and DNA sequencing.

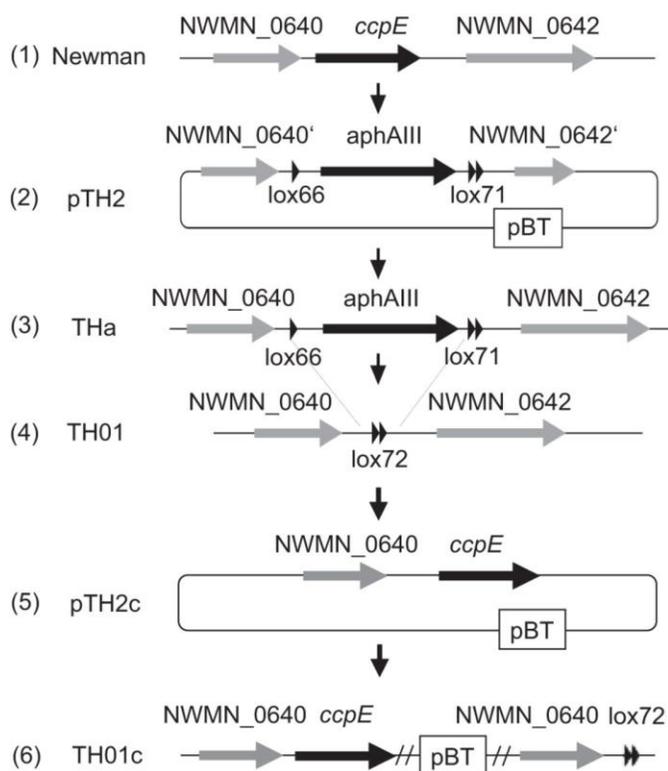


Figure 8.1 Schematic representation of the *ccpE* region of *S. aureus* and the strategies used to obtain the Δ *ccpE* mutant TH01 and the *cis*-complemented derivative TH01c. 1) genetic organization of the *S. aureus* Newman *ccpE* region. Open reading frames (arrowed boxes) and promoters (vertical arrows) are indicated. 2) a 3-kb fragment containing the *lox66-aphAIII-lox71* resistance cassette and the flanking regions of the *ccpE* gene was cloned into the vector pBT to generate plasmid pTH2. 3) plasmid pTH2 was electroporated into *S. aureus* strain RN4220 to obtain THa (RN4220 Δ *ccpE*::*lox66-aphAIII-lox71*), which was used as a donor for transducing the *aphAIII*-tagged *ccpE* deletion into other *S. aureus* strains. 4) a marker-free Δ *ccpE* mutant of *S. aureus* strain Newman was obtained by treating a Δ *ccpE*::*lox66-aphAIII-lox71* positive derivative of strain Newman with a Cre recombinase. Dotted lines between 3 and 4 indicate the region removed by Cre. 5) a 2-kb fragment covering the *ccpE* ORF and 1 kb of the upstream region including the NWMN_0640 ORF was cloned into the vector pBT to generate plasmid pTH2c. 6) a dam- and dcm-methylation free aliquot of plasmid pTH2c was directly electroporated into TH01 to obtain Th01c.

8.2.3 Construction of the *S. aureus ccpE* cis-complementation strain TH01c

For the *cis*-complementation of the *ccpE* mutation in strain TH01, a 2 kb fragment containing the wild-type *ccpE* allele and 1 kb of the upstream region including the NWMN_0640 open reading frame, was amplified by PCR from chromosomal DNA of *S. aureus* strain Newman using the primer-pair MBH225/MBH226. The resulting PCR product was digested with *Bam*HI/*Kpn*I, and subsequently cloned into *Bam*HI/*Kpn*I-digested vector pBT to generate the suicide plasmid pTH2c. The plasmid was transformed into and purified out of the *dam*- and *dcm*-negative *E. coli* strain SCS110 (Stratagene), and electroporated into *S. aureus* strain TH01 to obtain the *cis*-complementation strain TH01c (TH01 Δ *ccpE*::*pTH2c*). Restoration of the *ccpE* wild-type gene was verified by sequencing of the respective DNA fragment.

8.2.4 Construction of the *B. subtilis ccpC* promoter - *B. subtilis ccpC* and *B. subtilis ccpC* promoter - *ccpE* trans-complementation plasmids

Nucleotide sequences covering the *B. subtilis ccpC* promoter and gene were amplified by PCR from *B. subtilis* strain AF21 using the primer pairs MBH341/MBH255 and cloned into *Xba*I/*Kpn*I digested shuttle vector pBus1.¹⁸ Similarly, the *ccpE* gene from *S. aureus* strain Newman was amplified by PCR using primers MBH239/MBH226 and the *ccpE* fragment was digested with *Sal*I/*Kpn*I and cloned into pBus1 that was digested with the same restriction enzymes. The *ccpC* promoters P1 and P2 were amplified by PCR by using primer pair MBH341/MBH338 and *B. subtilis* strain AF21 DNA. The PCR product was digested with *Xba*I/*Sal*I and cloned into pre-digested pBus1 harboring the *S.*

aureus ccpE open reading frame. The resulting plasmids, pTH3 (*Bs ccpC* P1/2-*ccpC*) and pTH4 (*Bs ccpC* P1/2-*Sa ccpE* fusion), were used to transform competent cells of CJB9 or to electroporate RN4220. The resulting derivatives of RN4220 were then used as a donor for transducing it into *S. aureus* strain TH01.

8.2.5 Aconitase activity assay

Aliquots (2 mL) of bacteria were harvested after 3, 6, 9 and 12 h of growth in LB-L, and suspended in 850 μ L lysis buffer (90 mM Tris pH 8.0 and 100 μ M fluorocitrate). Bacteria were mechanically disrupted in a Fast Prep instrument (Qbiogene, Heidelberg, Germany) at a speed of 6.0 for 30 sec. The lysate was centrifuged for 1 min at 15,000 x g and 4 $^{\circ}$ C. Aconitase activity was assayed in the cell-free lysate as described previously.²⁰ One unit of aconitase activity was defined as the amount of enzyme necessary for a ΔA_{240} min⁻¹ of 0.0033.²¹ Total protein concentrations were determined according to the method described by Bradford.²²

8.2.6 Citrate synthase activity assay

Citrate synthase activity was determined with the Citrate Synthase Assay Kit purchased from Sigma-Aldrich (Taufkirchen, Germany) with slight modifications. Aliquots (2 mL) of bacteria were harvested after 3, 6, 9 and 12 h of growth in LB-L and suspended in CellLytic M Cell Lysis Reagent. Bacteria were mechanically disrupted in a Fast Prep instrument (Qbiogene) at a speed of 6.0 for 30 sec. The lysate was centrifuged for 1 min at 15,000 x g at 4 $^{\circ}$ C, and 10 μ L of the cell free lysate was mixed at room

temperature with 0.1 mM 5,5'-dithiobis-(2-nitrobenzoic acid) (DTNB) and 0.3 mM acetyl coenzyme A (CoA) in 1x Assay buffer (Sigma-Aldrich). The solution was mixed gently, and the absorbance of the reaction mixture at 412 nm (A_{412}) was followed for 1.5 min to obtain a background reading. After the addition of 0.5 mM oxaloacetate (OAA), the absorbance of the reaction mixture at A_{412} was monitored for an additional 1.5 min to detect the formation of 5-thio-2-nitrobenzoic acid (TNB). Citrate synthase activity units were calculated as micromoles of TNB produced (*i.e.*, CoA-SH released) per minute per milligram of total protein ($\mu\text{mol}/\text{min}/\text{mg}$). Total protein concentrations were determined according to the method described by Bradford.²²

8.2.7 Determination of citrate by GC/MS

Lyophilized cell pellets (10 mg dry weight) of Newman, TH01, and TH01c (each harvested after 8 h of growth in LB-L) were combined with 0.5 mL glass beads (Roth, Karlsruhe, Germany; 0.25 – 0.5 mm diameter) and 1 mL of deuterated methanol (MeOD_4) was added. This mixture was subjected to homogenizer lysis (FastPrep FP120, QBiogene, USA; 3 cycles each of 20 s at 6.5 m/s). After cell disruption, the cell debris and glass beads were separated by centrifugation (10 min at 5,000 rpm). The supernatant (0.5 mL) was dried under a stream of nitrogen. The residue was dissolved in 50 μL of methoxyamine hydrochloride (solution of 20 mg per mL pyridine) and reacted at 30 °C for 90 min. 50 μL of N-methyl-N-(trimethylsilyl) trifluoroacetamide with 1 % trimethylchlorosilane (Fluka, Germany) were added and the mixture was incubated at 37 °C for 30 min. GC/MS analyses were performed on a GC/MS Shimadzu QP 2010 Plus

(Shimadzu, Duisburg, Germany) equipped with a fused silica capillary column (Equity TM-5; 30 m × 0.25 mm, 0.25 μm film thickness; SUPELCO, Bellefonte, PA) and a QP-5000 mass selective detector (Shimadzu, Duisburg, Germany) working with electron impact ionization at 70 eV. An aliquot (1 μL) of the solution was injected in 1:10 split mode at an interface temperature of 250 °C and a helium inlet pressure of 76 kPa. The column was developed at 70 °C for 5 min and then with a temperature gradient of 5 °C/min to a final temperature of 310 °C that was held for 1 min. Data were collected using the LabSolutions software (Shimadzu, Duisburg, Germany). Compound peaks were assigned by comparison of the resulting mass spectra with those of a spectral library (NIST05, Shimadzu) and with data of injected reference samples under the same conditions.

8.2.8 Transcriptional analyses

For Northern blot experiments, overnight cultures of *S. aureus* were diluted to an A_{600} of 0.05 into fresh pre-warmed LB-L and grown at 37 °C and 230 rpm. Samples were removed from the culture at the indicated times and centrifuged at 9,000 × g and 4 °C for 2 min, the culture supernatants were discarded, and the cell pellets were snap frozen in liquid nitrogen. Total RNAs were isolated according to,²³ and blotting, hybridization, and labeling were performed as described previously.²⁴ Primer pairs MBH279/MBH280 and MBH281/MBH282 were used to generate digoxigenin-labeled NWMN_0640- and *ccpE*-specific probes by PCR labeling, respectively.

For the quantification of transcripts by real-time RT-PCR, RNA isolations and real-time RT-PCRs were carried out essentially as described in (17).²⁵ The obtained cDNA was used for real-time amplification with specific primers (Table 8.1) and 20 ng of cDNA per reaction. mRNA levels were normalized against the mRNA level of *gyrB*, which is constitutively expressed under the conditions analyzed.²⁶ The amounts of different transcripts were expressed as the *n*-fold difference relative to the control gene ($2^{-\Delta C_T}$, where ΔC_T represents the difference in threshold cycle between the target and control genes).

8.2.9 Determination of transcriptional start sites of *NMMN_0640* and *citB*

Strain Newman TF was grown for 3h as described and the bacteria were harvested by centrifugation at $8,000 \times g$ for 10 min. RNA isolation was carried out as described above. The transcriptional start points of *NMMN_0640* and *citB* were determined by rapid amplification of cDNA ends essentially as described by (19).²⁷ The reverse transcription of 8.2 μL of the ligated RNA was carried out with *NMWN_0640* or *citB* specific primers (Table 8.1) with the High Capacity cDNA Reverse Transcription Kit (Life Technologies) according to the manufacturer's instructions. 2 μL of the cDNA was amplified by PCR using a primer complementary to the RNA adapter sequence, and a specific primer for *NMWN_0640* or *citB* (Table 8.1) closer to the 5'-end than the primers used for the reverse transcription. Finally, the PCR products were sequenced.

8.2.10 Antibody production and Immunoblotting

Polyclonal anti-CcpE antibodies were raised by injecting 500 µg of the His-tagged recombinant CcpE into rabbits (Eurogentec, Liege, Belgium). The resulting crude antisera were purified against the immobilized CcpE antigen. For the determination of CcpE, cytoplasmic protein extracts were isolated from *S. aureus* cell cultures grown for 3, 6, 9 and 12 h in LB-L at 37 °C as described previously,²⁸ and protein fractions (20 µg/lane) were separated using SDS-PAGE, blotted onto a nitrocellulose membrane, and subjected to Western blot analysis using the antigen-purified polyclonal anti-CcpE antiserum.

8.2.11 Electrophoretic Mobility Shift Assays

The DNA probes for electrophoretic mobility shift assays (EMSAs) were generated by PCR using *S. aureus* strain Newman chromosomal DNA as a template, and primer pairs (listed in Table 8.1) that amplified the DNA regions preceding the *ccpE*, *citB*, *citZ*, and NWMN_0640 ORFs. The 5'-ends of the double-stranded PCR products were labeled using [γ -³²P] ATP and T4 polynucleotide kinase. A typical assay mixture contained (in 20 µL) 10 mM Tris-HCl, pH 7.5, 50 mM NaCl, 1 mM EDTA, 1 mM DTT, 0.1 µg of nonspecific competitor (poly(dI-dC)), 5% (v/v) glycerol, radioactive DNA probe (2000 cpm mL⁻¹), and various amounts (0, 15, 65, 130, and 200 nM) of purified CcpE. After 30 min of incubation at room temperature, 20 µL of this mixture was loaded into a native 4% (w/v) polyacrylamide Tris borate-EDTA Ready Gel (Bio-Rad) and electrophoresed in 1% Tris borate-EDTA (v/v) buffer for 1 h at 100 V cm⁻¹. Radioactive species were detected by autoradiography using direct exposure to films.

8.2.12 NMR Sample preparation

S. aureus strains were grown in filter-sterilized Luria-Bertani Lennox (LB-L) medium (BD, Heidelberg, Germany) supplemented with 0.1% glucose (Sigma-Aldrich). Bacteria from overnight cultures were diluted into pre-warmed medium to an optical density at 600 nm (OD_{600}) of 0.05. All bacterial cultures were incubated at 37°C under aerated conditions at 225 rpm with a flask-to-medium ratio of 10:1. For two-dimensional (2D) ^1H , ^{13}C HSQC analysis LB-L medium containing 0.1% $^{13}\text{C}_6$ -glucose (Cambridge Isotope Laboratories) was used for main cultures. NMR samples for intracellular metabolite analysis were prepared from independent cultures in exponential (3 h) and post-exponential (8 h) growth phase. The extraction of the metabolome from cell lysates followed our previously published protocols.^{29,30}

8.2.13 NMR data collection, analysis and interpretation

Followed as previously published.^{29,30}

8.2.14 Statistical analyses

Statistical significance was assessed using Student's *t*-test or Mann–Whitney U test. *P* values < 0.05 were considered significant.

8.3 Results

8.3.1 Identification of potential carbon metabolism-affecting factors in *S. aureus* strain Newman

Newman

A bioinformatic comparison of the genomes of *B. subtilis* isolate 168 (Genbank acc. no. AL009126.3) and *S. aureus* strain Newman (acc. no. AP009351.1) suggested the strain Newman genome might have genes whose products have been associated with carbon catabolite repression (CCR) in *B. subtilis* (reviewed in reference 1). One uncharacterized gene was NWMN_0641 (renamed here as *ccpE*), which shared 61% similarity and 35% identity to the *B. subtilis* citrate-responsive regulator CcpC, and like CcpC it was predicted to encode for a putative transcriptional regulator of the LysR family. These observations and the fact that CcpC is a regulator of tricarboxylic acid (TCA) cycle genes *citZ* (encoding citrate synthase) and *citB* (encoding aconitase)⁶, led us to examine the function of CcpE in *S. aureus*.

8.3.2 Transcriptional organization of the *ccpE* locus in *S. aureus* strain Newman

Our microarray analysis of σ^B -mediated transcriptional changes suggested that *ccpE* might form a bi-cistronic operon with the open reading frame (ORF) NWMN_0640.³¹ Further bioinformatic analysis of this region suggested that both genes were likely controlled by a promoter located upstream of NWMN_0640. To test this suggestion, we assessed the transcription of NWMN_0640 and *ccpE* by Northern blotting (Figure 8.1a). This Northern blot analysis revealed *ccpE*-specific transcripts with sizes of approximately 4.1-, 2.9-, 2.3-, and 1.5-kb. The 2.9- and 1.5-kb transcripts migrated on the gel to the same extent as the 16S- and 23S rRNAs; hence, they might represent

degradation products of higher molecular weight transcripts. That being said, all transcripts were detectable throughout the growth cycle and the transcription profiles of *ccpE* and NWMN_0640 were nearly identical. Based on these results, and on the genetic organization of the *ccpE* locus (Figure 8.1b), we hypothesized that all transcripts originate from a single promoter in front of NWMN_0640, which would give rise to a NWMN_0640/*ccpE* encoding 2.3-kb transcript, and a larger NWMN_0640/*ccpE* /NWMN_0642/3 encoding 4.1-kb transcript.

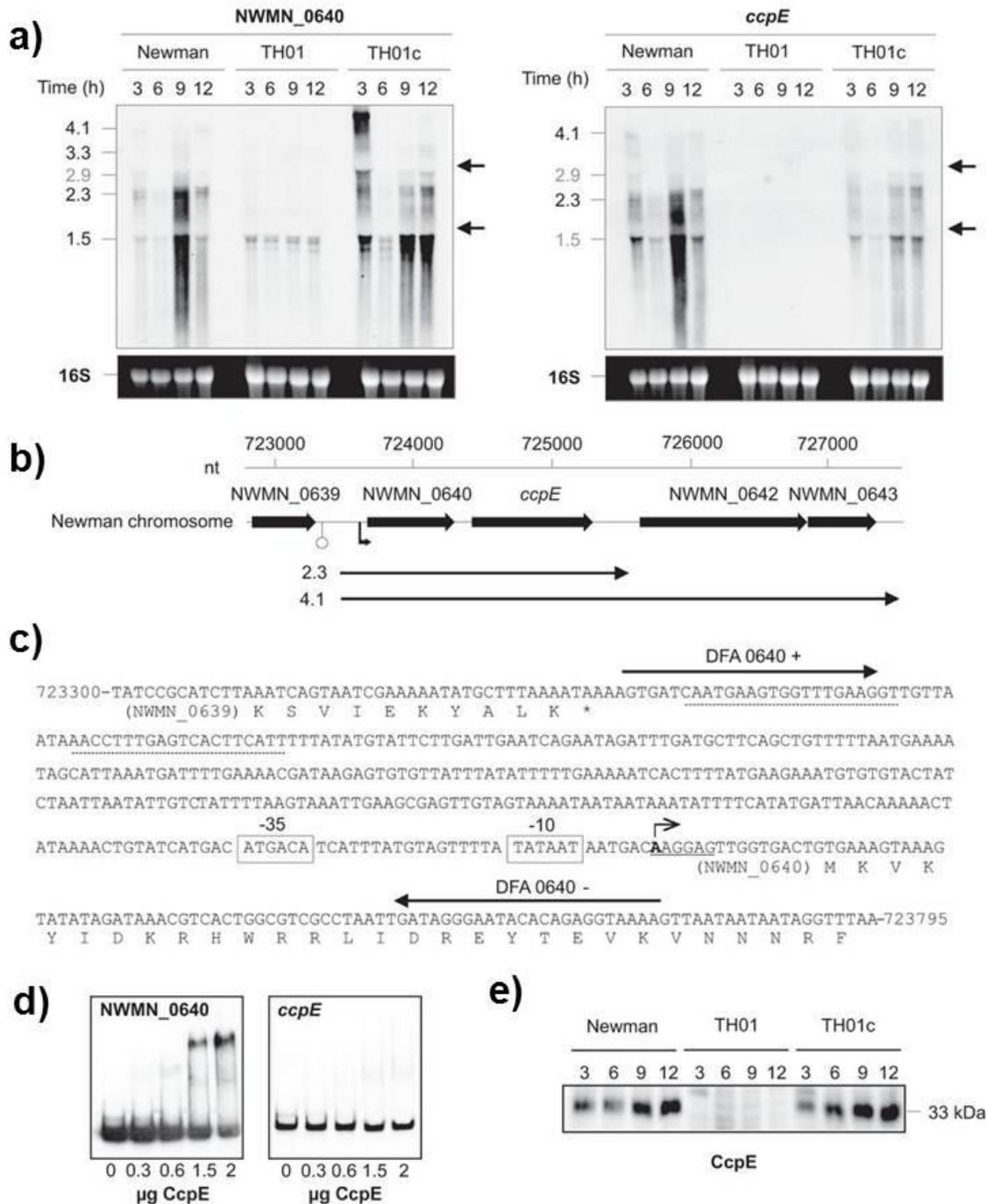


Figure 8.2 Transcriptional organization of the *ccpE* locus of *S. aureus* Newman
 A, Northern blot of NWMN_0640 and *ccpE* transcriptions in strains Newman, TH01 (Δ *ccpE*), and the complemented TH01c during growth in LB-L. Approximate transcript sizes are indicated on the *left*. Ethidium bromide-stained 16 S rRNA are presented to

indicate equivalent RNA loading. *B*, schematic representation of the *ccpE* region of *S. aureus*. Proposed ORFs, promoters, terminators, and transcripts identified by Northern analyses are indicated. ORF notations and nucleotide (*nt*) numbers correspond to those of the respective genomic regions of strain Newman (GenBank accession number AP009351.1). *C*, nucleotide sequence of the region preceding the NWMN_0640 open reading frame. The transcriptional start point identified by RACE (*bent arrow*), putative -35 and -10 boxes (*boxes*), terminator sequences (*hatched lines*), and primers used to amplify the NWMN_0640-*ccpE* promoter fragment used in EMSA (*horizontal arrows*) are indicated. *D*, binding activity of CcpE to the DNA regions preceding the NWMN_0640 and *ccpE* ORFs. The PCR-amplified DNA fragments were radioactively labeled and incubated with the amount of purified CcpE indicated. The results are representative of at least two independent experiments. *E*, Western blot analysis of CcpE in cytosolic extracts of strains Newman, TH01, and TH01c cells grown in LB-L to the time points indicated. The results are representative of at least two independent experiments.

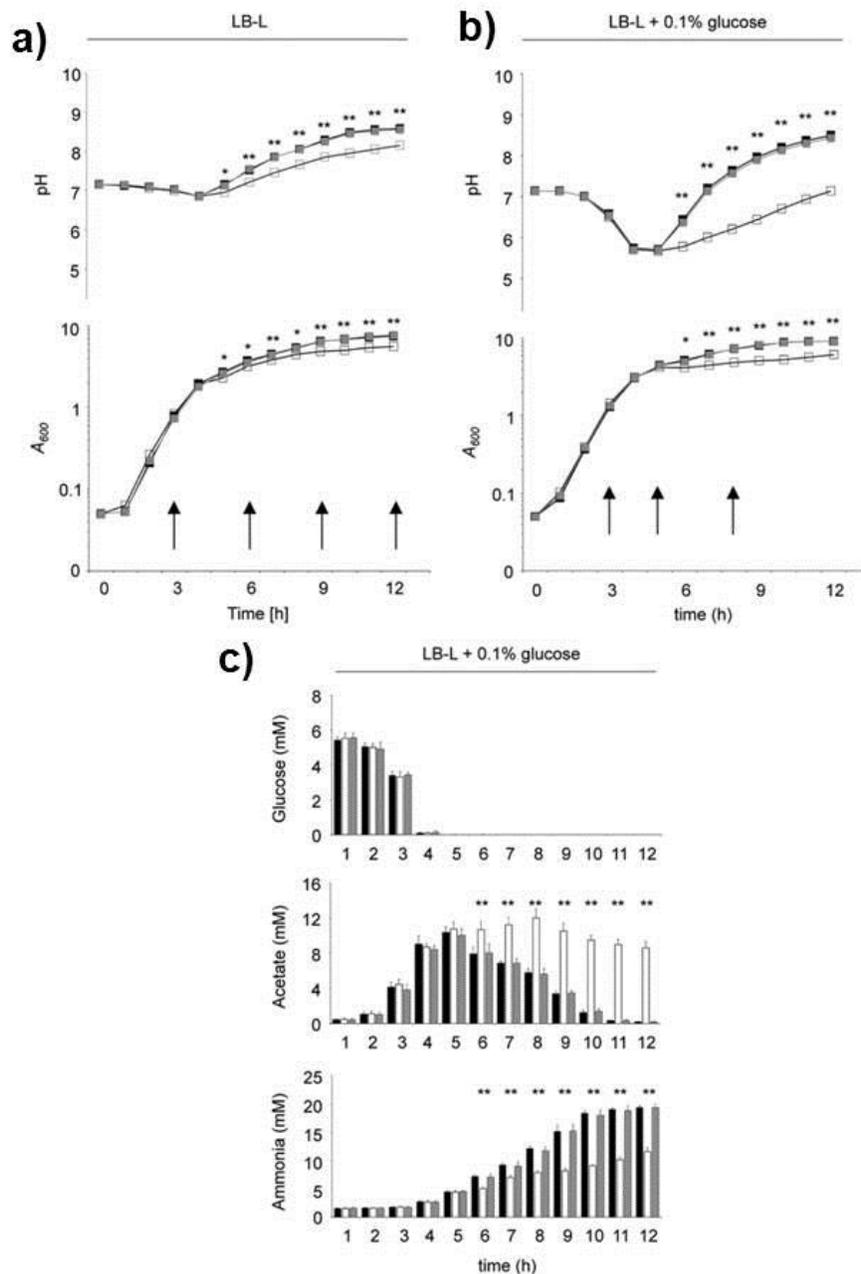


Figure 8.3 Growth characteristics of *S. aureus* strain Newman (black symbols), TH01 (white symbols), and TH01c (gray symbols). Culture media pH and A_{600} values during growth in LB-L medium (A) and LB-L supplemented with 0.1% glucose (B). Time points of sampling for downstream applications (enzyme assays, qRT-PCRs) are indicated by arrows. The data presented are mean \pm S.D. of at least three independent experiments.

Student's *t* test *, $p < 0.05$; **, $p < 0.01$. *C*, glucose, acetate, and ammonia concentrations in the culture supernatants of Newman (*black bars*), TH01 (*white bars*), and TH01c (*gray bars*) cell cultures grown in LB-L + 0.1% glucose. The data presented are mean \pm S.D. of at least three independent experiments done in duplicate. Mann-Whitney *U* test. **, $p < 0.01$.

To identify the transcriptional start point (TSP) of NWMN_0640, we performed a 5'RACE experiment which identified a TSP 14 bp upstream of the proposed start point of the NWMN_0640 open reading frame, and immediately upstream of the putative ribosomal binding site of this ORF (Figure 8.1c). The TSP is preceded by the nucleotide sequence ATGACA-17-TATAAT that strongly matched with the -35 and -10 hexamers identified in the promoters of genes shown to be transcribed by *S. aureus* σ^A containing RNA-polymerase.^{32,33}

Since many LysR-type of regulators display autocrine regulation of their own transcription,³⁴ we determined the capacity of CcpE to bind to its own promoter. To do this, purified CcpE was used in electrophoretic mobility shift assays with a PCR probe covering the genomic region preceding the NWMN_0640 ORF (Figure 8.1d). As expected, a clear, dose-dependent shift with CcpE and the radioactive-labeled NWMN_0640 promoter probe fragment was observed, suggesting that CcpE affects its own transcription. This DNA binding activity was specific, as an EMSA using a DNA probe covering the region preceding the *ccpE* ORF did not cause a mobility shift. Taken together, our data suggest that *ccpE* is transcribed in a bi-cistronic message using a

promoter upstream of *NWMN_0640*, and that this promoter is subjected to autocrine regulation by CcpE.

8.3.3 Inactivation of *ccpE* in *S. aureus* strain Newman

Using the *cre-lox* based deletion system described by Leibig and colleagues,¹⁷ we created a marker-less deletion of *ccpE* in *S. aureus* strain Newman, giving rise to strain TH01. In addition, a *cis*-complemented derivative was created by integrating the *NWMN_0640/ccpE* containing suicide plasmid, pTH2c, into the deletion site of strain TH01, resulting in strain TH01c. Sequencing of the respective genome regions confirmed the mutations occurred as expected. Further, Northern blot analysis qualitatively confirmed that strain TH01 did not produce a *ccpE*-specific transcript, while strain TH01c expressed *ccpE* at a level comparable to that of the wild-type strain (Figure 8.1a). Probing total RNA from TH01 and TH01c with a *NWMN_0640*-specific probe also confirmed the deletion of *ccpE* in TH01, as we detected in strain TH01 *NWMN_0640*-specific transcripts of approximately 3.3- and 1.5-kb. These mRNA sizes are consistent with the deletion of 800 bp of the *ccpE* locus in this mutant. Importantly, complementation restored all *NWMN_0640*-specific transcripts found in the wild-type strain and also transcripts that were identified in TH01. This latter result was expected due to our complementation strategy that created a duplication of the *NWMN_0640* gene due to the insertion of the suicide plasmid pTH2c in the *NWMN_0640* open reading frame. Taken together, these data confirmed that the correct gene was inactivated and that wild-type transcription profiles could be restored by complementation.

To assess the production of CcpE in *S. aureus*, rabbit polyclonal antibodies were generated against CcpE and used in Western blot analyses of cytosolic protein fractions from strains Newman, TH01, and TH01c grown in LB-L (Figure 8.2). In these anti-CcpE Western blots, 33 kDa bands in both the wild-type strain and complemented strain TH01c were observed throughout the growth cycle and, as expected, this band was absent in the *ccpE* mutant strain TH01. These data were consistent with the Northern data (Figure 8.1a) and indicated that CcpE accumulated in the cytosols of the wild-type and TH01c strains during the later stages of growth.

8.3.4 CcpE affects the in vitro growth yield of *S. aureus*

To determine if *ccpE* inactivation affected the physiology of *S. aureus*, we assessed the growth and the culture medium pH profiles of strains Newman, TH01, and TH01c in LB-L medium in absence or presence of 0.1% supplemental glucose (Figure 8.3a). In LB-L medium lacking glucose, strain TH01 displayed a slight decrease in the growth yield relative to the wild-type and TH01c strains (Figure 8.3a). Additionally, the pH of the culture medium for strain TH01 was less alkaline than for the wild-type and TH01c strains, indicating that the *ccpE* mutant was either impaired in amino acid catabolism or that the accumulation or depletion of organic acids has changed. Supplementation of LB-L with 0.1% glucose increased the overall growth yields of all three strains relative to cultures grown in LB-L, but it did not alter the growth rate (Figure 8.3b). The similarity of the growth rates but the differences in growth yields suggested that exponential growth phase metabolism was similar but that post-

exponential growth phase metabolism was altered. Consistent with this suggestion, the pH profiles and glucose, acetate, and ammonia accumulation in the medium were nearly identical up until the post-exponential growth phase (Figure 8.3). In the post-exponential growth phase, the culture medium of strain Newman and TH01c began to alkalinize as acetate was extracted from the medium for catabolism via the TCA cycle, and ammonia accumulated due to the deamination of amino acids (Figure 8.3c). In contrast to strains Newman and TH01c, alkalization of the culture medium from strain TH01 was very slow, as both acetate catabolism and ammonia accumulation were repressed. This reduced ability to catabolize acetate and amino acids was also reflected in the decreased biomass generation and final growth yield. As acetate catabolism and amino acid catabolism were repressed, these data suggested that *ccpE* inactivation inhibited/decreased TCA cycle activity.

8.3.5 Deletion of *ccpE* alters the *S. aureus* metabolome

The deletion of *ccpE* changed the accumulation and depletion of organic acids in the culture medium and the pH profile (Figure 8.3), suggesting that *ccpE* inactivation caused significant metabolic changes. To test this hypothesis, strains Newman, TH01, and TH01c were grown in LB-L medium containing 0.1% ^{13}C -glucose and the metabolomes were analyzed by NMR spectroscopy. The metabolomic analysis of strains Newman, TH01, and TH01c grown in LB-L medium containing 0.1% ^{13}C -glucose provided a metabolic snapshot of the exponential (3 h) and post-exponential (8 h) growth phases (Figure 8.4). Consistent with growth and pH profiles for the strains (Figure 8.3b),

an exponential growth phase PCA plot (Figure 8.4a) revealed no significant differences in the metabolomes of strains Newman, TH01, and TH01c. In contrast to the exponential growth phase, *ccpE* inactivation had a major effect on post-exponential growth phase metabolism (Figure 8.4b-e). In particular, carbon flow through the TCA cycle was reduced as evidenced by the accumulation of citrate and the shunting of pyruvate into fermentative pathways. The accumulation of citrate in strain TH01 was also confirmed in bacteria cultivated in LB-L without glucose, using gas chromatography-mass spectrometry (GC-MS) (Figure 8.5). The metabolic block in the TCA cycle created by *ccpE* inactivation led to a reduction in the intracellular concentrations of glutamate and glutamine, which affects the availability of nitrogen donors. The decreased ammonia assimilation via glutamate to glutamine may be the cause for an increased concentration of asparagine, which is generated by amination of aspartate via asparaginase. In summary, we can conclude that CcpE is involved in regulating carbon flow through the TCA cycle, independent of glucose or its catabolic products.

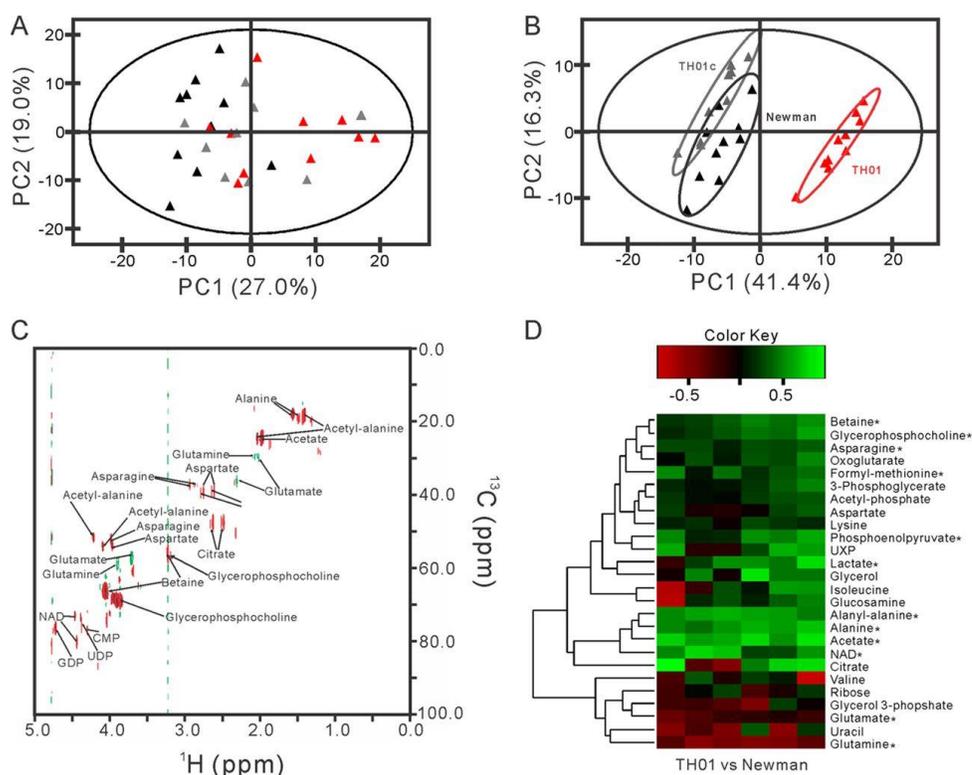


Figure 8.4 Effect of the *ccpE* deletion on the metabolome of *S. aureus* strain Newman. A and B, two-dimensional PCA scores plots generated from the entire one-dimensional ^1H NMR spectra of the exponential growth phase (3 h) metabolomes (A), and post-exponential growth phase (8 h) metabolomes (B) of strains Newman (*black triangles*), TH01 (*red triangles*), and TH01c (*gray triangles*). C, two-dimensional ^1H , ^{13}C -HSQC difference spectrum showing the post-exponential growth phase differences between strains Newman (*green*) and TH01 (*red*). D, a heat map of six independent replicates showing the major differences in metabolites produced in post-exponential growth phase by strain TH01 in relationship to strain Newman. Relative concentration changes were plotted as outlined under “Experimental Procedures,” using a color-scale from -1 (*red*) to 1 (*green*). The dendrogram depicts hierarchical clustering of the metabolite concentration changes, where metabolites exhibiting the same relative trend and magnitude changes across all replicates are clustered together ($n = 6$). A Student's *t* test was used to determine the statistical significance of each metabolites change between the TH01 and Newman strains (*, $p < 0.05$).

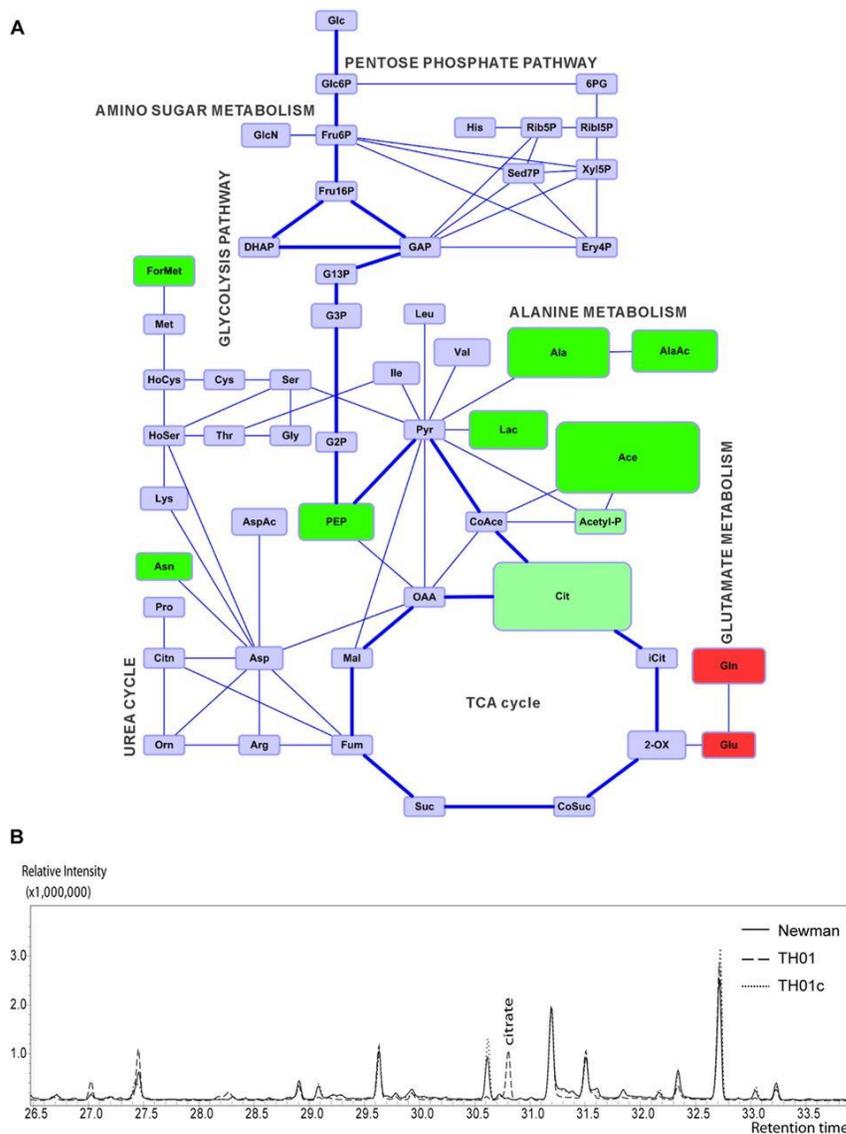


Figure 8.5 *CcpE*-dependent changes in the metabolome of *S. aureus*. *A*, a cytoscape map showing the changes in the central metabolism associated with *ccpE* inactivation. *B*, overlays of representative GC/MS profiles of Newman, TH01, and TH01c whole cell extracts of 8-h LB-L cultures. The elevated citrate peak observed with TH01 is indicated.

8.3.6 Inactivation of *ccpE* affects transcription of TCA cycle genes and its activity

The CcpE open reading frame in *S. aureus* was identified based on its homology to CcpC in *B. subtilis*. CcpC is involved in the regulation of the TCA cycle genes *citZ* and *citB*;^{5,6,8} hence, we hypothesized that CcpE might regulate the same TCA cycle genes in *S. aureus*. To test this assumption, transcription of the TCA cycle genes *citZ* and *citB* was assessed by qRT-PCR in strains Newman, TH01, and TH01c grown in LB-L medium (Figure 8.6). In the wild-type strain Newman, *citZ* and *citB* mRNAs were at the highest levels just prior to the post-exponential growth phase (Figure 8.6a). Inactivation of *ccpE* decreased transcription of both *citB* and *citZ*; however, the effect on *citB* was more dramatic. Complementation of the *ccpE* mutation restored *citZ* and *citB* transcription to levels similar to that in the wild-type strain. Thus, in order to determine if the differences in *citB* and *citZ* transcription were reflected in enzymatic activity changes, the activities of citrate synthase and aconitase were measured in the wild-type, TH01, and TH01c strains as well (Figure 8.6b). Consistent with the transcriptional data, citrate synthase activity was decreased in strain TH01 at 3 h of growth only, while aconitase activity was significantly decreased in TH01 cells in all growth phases. Also consistent with the transcriptional data, complementation of the *ccpE* mutation restored citrate synthase and aconitase enzymatic activities to those in the wild-type strain, strongly suggesting that CcpE is a major positive transcriptional regulator of *citB* in *S. aureus*.

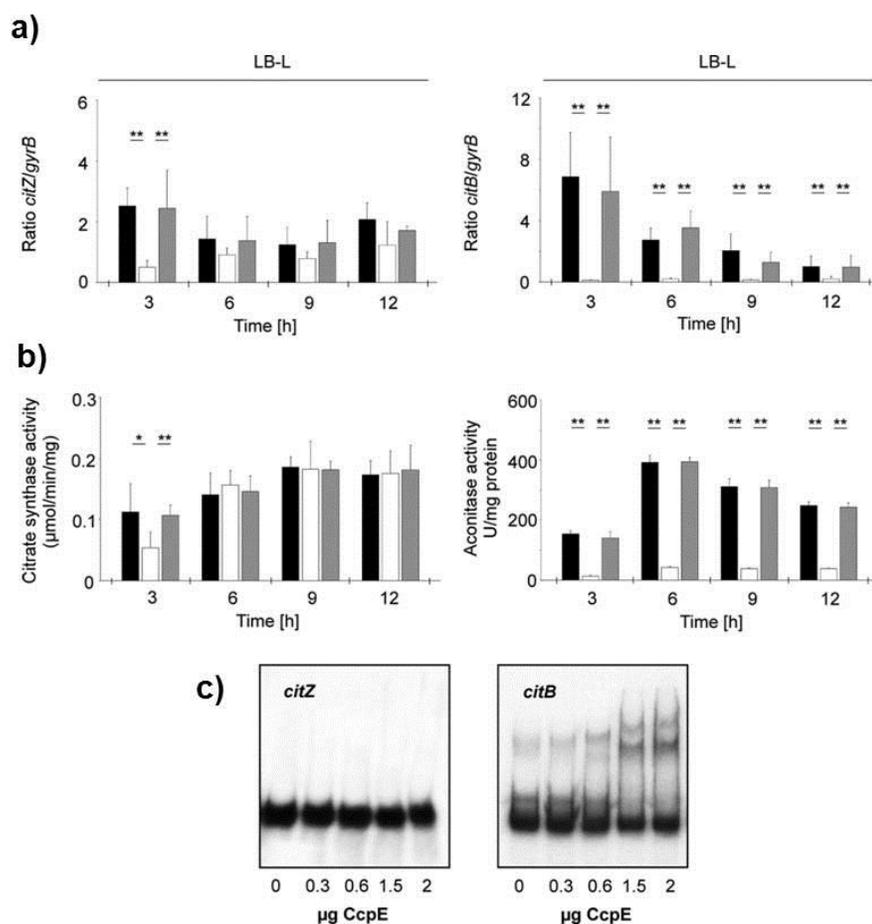


Figure 8.6 Transcription and activity of *citZ* and *citB* in strain Newman and its derivatives. *A*, quantitative transcript analysis of *citZ* and *citB* by qRT-PCR of strains Newman (*black bars*), TH01 (*white bars*), and TH01c (*gray bars*) during growth in LB-L. mRNA levels are expressed relative to gyrase B (in numbers of copies per copy of *gyrB*). The data presented are mean \pm S.D. of four independent experiments each determined in duplicate. *B*, enzyme activities of citrate synthase, and aconitase of strains Newman (*black bars*), TH01 (*white bars*), and TH01c (*gray bars*) during growth in LB-L. The data are presented as the mean \pm S.D. of three independent experiments done in duplicate. Mann-Whitney *U* test *, $p < 0.05$; **, $p < 0.01$. *C*, EMSA using purified CcpE and the DNA probes generated from the *citB* and *citZ* promoters. The PCR-amplified DNA fragments were radioactively labeled and incubated with the amount of purified CcpE indicated. The results are representative of at least two independent experiments.

8.3.7 CcpE binds to the *citB* promoter

To assess whether CcpE might directly regulate transcription of *citB* and *citZ* by binding to the respective promoters, we performed EMSAs using the *citZ* and *citB* promoters as probes (Figure 8.6c). The probe generated from the *citB* promoter shifted with CcpE in a dose-dependent manner; however, the *citZ* promoter region was not shifted by CcpE, suggesting that CcpE directly controls the expression of *citB* but not *citZ*. The similarity of CcpE to CcpC and the fact that citrate accumulated when *ccpE* was inactivated, led us to determine if citrate would influence CcpE DNA binding activity. In contrast to *B. subtilis* CcpC, the binding activity of CcpE was independent of the concentration of citrate in the binding buffer. Similarly, the CcpE binding was independent of NAD⁺ and NADH. These data suggest that CcpE is not a functional ortholog of CcpC, but a newly described TCA cycle regulator.

8.3.8 CcpE is not a functional homolog of CcpC

In *B. subtilis*, CcpC regulates transcription of TCA cycle genes in response to changes in citric acid, and inactivation of *ccpC* increases the exponential growth phase transcription of *citB*.⁶ Although *ccpE* inactivation decreases *citB* transcription and its DNA binding properties are not altered by citric acid; out of an abundance of caution, we assessed whether the *S. aureus ccpE* gene could complement a *B. subtilis ccpC* mutant and whether the *B. subtilis ccpC* gene could complement our *S. aureus ccpE* mutant. To do this, plasmids having *ccpC* under the control of its native promoter (pTH3) and a fusion plasmid harboring the *B. subtilis ccpC* P1 and P2 promoters fused to the *S. aureus*

ccpE ORF (pTH4) were constructed. These plasmids were transformed into TH01 and the *B. subtilis ccpC* mutant CJB9,⁶ respectively, and the final transformants were tested for CcpE by Western blotting and *citB*, *ccpC*, and *ccpE* transcription using qRT-PCR (Figure 8.7). As expected, *citB* transcription was increased in the *B. subtilis ccpC* mutant strain CJB9 when grown in TSS minimal medium supplemented with 0.2 % (w/v) glutamine and 0.5 % (w/v) glucose. Complementation of strain CJB9 with *ccpC* under the control of its native promoter, plasmid pTH3, restored *citB* transcription to wild-type levels. In contrast, transforming strain CJB9 with plasmid pTH4, harboring the *S. aureus ccpE* under the control of the *B. subtilis ccpC* P1 and P2 promoters, failed to revert *citB* transcription to wild-type levels. To exclude that the inability of CcpE to restore wild-type *citB* mRNA levels was due to the fact that it was not transcribed or translated, we assessed the *ccpE* transcription using qRT-PCR and CcpE translation by Western blotting. In strain CJB9 containing plasmid pTH4, *ccpE* was strongly transcribed and CcpE was produced in large quantities (Figure 8.7a), demonstrating that the inability of *ccpE* to complement a *ccpC* mutation was neither due to a failure of transcription nor translation. Similarly, transforming strain TH01 with plasmids pTH3 (containing *ccpC*) or pTH4 (containing *ccpE*) increased *citB* transcription to wild-type levels with *ccpE* but not *ccpC* (Figure 8.7b). Taken together, these data demonstrate that CcpE is not a functional homolog of CcpC.

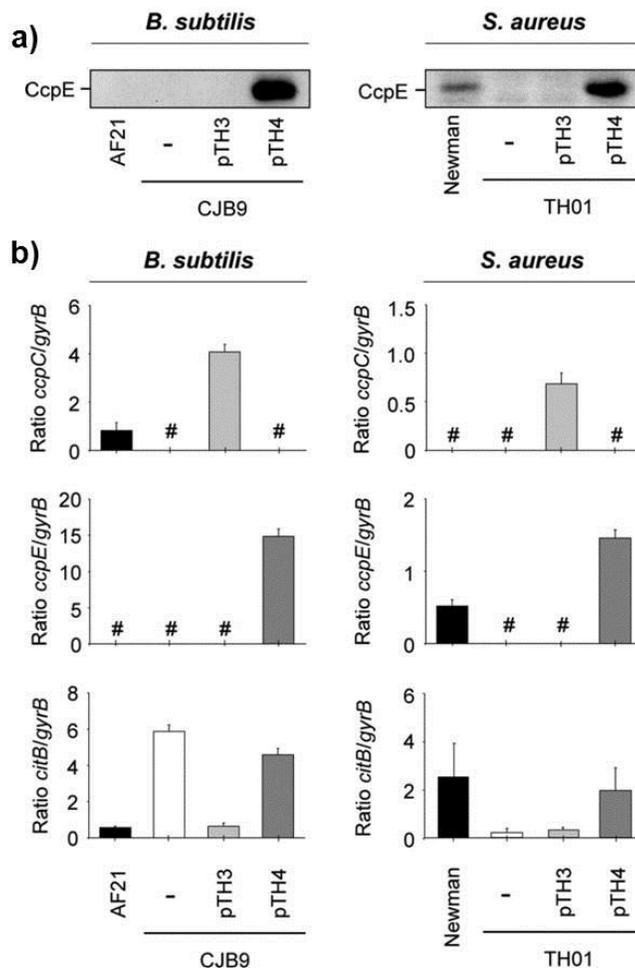


Figure 8.7 Trans-complementation of *B. subtilis* *ccpC* and *S. aureus* *ccpE* mutants. *A*, Western blot analysis of CcpE in cytosolic extracts of *B. subtilis* strains AF21, CJB9 (AF21::*ccpC*), CJB9 + pTH3 (*ccpC* under the control of its native promoters), CJB9 + pTH4 (*ccpC* P1/2-*ccpE* fusion), *S. aureus* strains Newman, TH01, TH01 + pTH3, and TH01 + pTH4 after growth in LB-L for 6 h. The results are representative of at least two independent experiments. *B*, quantitative transcript analyses of *ccpC*, *ccpE*, and *citB* by qRT-PCR of *B. subtilis* derivatives grown in TSS minimal medium supplemented with 0.2% (w/v) glutamine and 0.5% (w/v) glucose, and *S. aureus* derivatives grown in LB-L to the mid-exponential growth phase. The data presented are mean \pm S.D. of four independent experiments each determined in duplicate. #, transcripts below detection limits.

8.4 Discussion

Central metabolism provides *S. aureus* with the thirteen biosynthetic intermediates from which it derives all macromolecules; hence, central metabolism is critically important for growth and survival.³⁵ The TCA cycle is one component of central metabolism that provides the bacterium with energy, reducing potential and three of the thirteen biosynthetic intermediates. When these 13 biosynthetic intermediates, or the amino acids, nucleic acids, fatty acids, *etc.* that these intermediates produce, are exogenously available, TCA cycle activity is largely repressed.³⁶ In the presence of readily catabolizable carbohydrates, such as glucose, transcription of TCA cycle genes is repressed by CcpA.³⁷ Similarly, transcription of TCA cycle genes is repressed by CodY, a highly conserved Gram-positive repressor that responds to intracellular concentrations of branched-chain amino acids (BCAA) and GTP.¹² When nutrients become growth limiting, transcription of TCA cycle genes is de-repressed and TCA cycle activity dramatically increases, allowing for the utilization of incompletely oxidized organic acids that accumulated in the culture medium.³⁸ This utilization of organic acids through the TCA cycle allows *S. aureus* to generate biosynthetic intermediates needed for precursor and macromolecular synthesis by shunting carbon into gluconeogenesis via PEP. In other words, the TCA cycle is an important metabolic pathway that facilitates *S. aureus* adaption to a nutrient-limited environment. The importance of the TCA cycle to the success of *S. aureus* as a pathogen is illustrated by mutagenesis studies that identified TCA cycle mutants to be attenuated in different mouse models^{39,40} and a *Caenorhabditis elegans*-killing model.^{41,42} More recently, the significance of TCA cycle activity during

the course of infection was also suggested by a study from Chaffin and colleagues,⁴³ where it was observed that *citB* transcription increased over time in a mouse pneumonia model.

The importance of the TCA cycle in metabolism, survival, and virulence factor synthesis, led us to search the *S. aureus* genomic DNA sequence for TCA cycle regulators that are present in *B. subtilis*. One such regulator that is present in *B. subtilis* but undescribed in *S. aureus* is CcpC. We identified a homologue of CcpC in the *S. aureus* genome, which we named CcpE. At the amino acid level, CcpE is similar to the *B. subtilis* CcpC; however, the similarities end there. Deletion of *ccpE* in *S. aureus* dramatically decreased transcription of *citB* and to a lesser extent *citZ*. This decreased transcription resulted in decreased TCA cycle activity, causing a metabolic block in the TCA cycle that increased the intracellular citrate concentration. The increased citrate accumulation and the inability to effectively fully oxidize carbohydrates likely caused the decreased growth yield. These data are in stark contrast to that of CcpC where inactivation of *ccpC* in *B. subtilis* de-represses *citB* and *citZ* transcription.⁶ Also unlike CcpC, the DNA binding activity of CcpE is not dependent upon the concentration of citrate. In summary, CcpE is a major positive regulator of TCA cycle activity that binds DNA independent of the citrate concentration.

TCA cycle activity has been also associated with virulence in staphylococci. In *S. aureus*, TCA cycle activity was found to be critical for the elaboration of a capsule.⁴⁴ In contrast, TCA cycle activity negatively affects synthesis of polysaccharide intercellular adhesin and biofilm formation.⁴⁵ Because CcpE increases TCA cycle activity in *S.*

aureus, it is likely that CcpE activity will affect virulence determinant biosynthesis and pathogenesis. Experiments are ongoing to address this question. Moreover, another unresolved question is what metabolite/co-factors alter the DNA binding properties of CcpE. We know some that do not affect CcpE binding to DNA; namely, citrate and NAD⁺/NADH. Lastly, we are in the process of identifying the DNA binding site for CcpE.

8.5 Reference

- (1) Fujita, Y. *Biosci Biotechnol Biochem* **2009**, *73*, 245.
- (2) Deutscher, J. *Curr Opin Microbiol* **2008**, *11*, 87.
- (3) Singh, K. D.; Schmalisch, M. H.; Stulke, J.; Gorke, B. *J Bacteriol* **2008**, *190*, 7275.
- (4) Lorca, G. L.; Chung, Y. J.; Barabote, R. D.; Weyler, W.; Schilling, C. H.; Saier, M. H., Jr. *J Bacteriol* **2005**, *187*, 7826.
- (5) Kim, H. J.; Kim, S. I.; Ratnayake-Lecamwasam, M.; Tachikawa, K.; Sonenshein, A. L.; Strauch, M. *J Bacteriol* **2003**, *185*, 1672.
- (6) Jourlin-Castelli, C.; Mani, N.; Nakano, M. M.; Sonenshein, A. L. *J Mol Biol* **2000**, *295*, 865.
- (7) Servant, P.; Le Coq, D.; Aymerich, S. *Mol Microbiol* **2005**, *55*, 1435.
- (8) Jin, S.; Sonenshein, A. L. *J Bacteriol* **1994**, *176*, 4680.
- (9) Galinier, A.; Haiech, J.; Kilhoffer, M. C.; Jaquinod, M.; Stulke, J.; Deutscher, J.; Martin-Verstraete, I. *Proc Natl Acad Sci U S A* **1997**, *94*, 8439.

- (10) Shivers, R. P.; Dineen, S. S.; Sonenshein, A. L. *Mol Microbiol* **2006**, *62*, 811.
- (11) Fiegler, H.; Bassias, J.; Jankovic, I.; Bruckner, R. *J Bacteriol* **1999**, *181*, 4929.
- (12) Somerville, G. A.; Proctor, R. A. *Microbiol Mol Biol Rev* **2009**, *73*, 233.
- (13) Kreiswirth, B. N.; Lofdahl, S.; Betley, M. J.; O'Reilly, M.; Schlievert, P. M.; Bergdoll, M. S.; Novick, R. P. *Nature* **1983**, *305*, 709.
- (14) Duthie, E. S. *J Gen Microbiol* **1952**, *7*, 320.
- (15) Fouet, A.; Sonenshein, A. L. *J Bacteriol* **1990**, *172*, 835.
- (16) Giachino, P.; Engelmann, S.; Bischoff, M. *J Bacteriol* **2001**, *183*, 1843.
- (17) Leibig, M.; Krismer, B.; Kolb, M.; Friede, A.; Gotz, F.; Bertram, R. *Appl Environ Microbiol* **2008**, *74*, 1316.
- (18) Rossi, J.; Bischoff, M.; Wada, A.; Berger-Bachi, B. *Antimicrob Agents Chemother* **2003**, *47*, 2558.
- (19) Geisbrecht, B. V.; Bouyain, S.; Pop, M. *Protein Expr Purif* **2006**, *46*, 23.
- (20) Kennedy, M. C.; Emptage, M. H.; Dreyer, J. L.; Beinert, H. *J Biol Chem* **1983**, *258*, 11098.
- (21) Baughn, A. D.; Malamy, M. H. *Proc Natl Acad Sci U S A* **2002**, *99*, 4662.
- (22) Bradford, M. M. *Anal Biochem* **1976**, *72*, 248.
- (23) Cheung, A. L.; Eberhardt, K. J.; Fischetti, V. A. *Anal Biochem* **1994**, *222*, 511.

- (24) McCallum, N.; Karauzum, H.; Getzmann, R.; Bischoff, M.; Majcherczyk, P.; Berger-Bachi, B.; Landmann, R. *Antimicrob Agents Chemother* **2006**, *50*, 2352.
- (25) Chatterjee, I.; Becker, P.; Grundmeier, M.; Bischoff, M.; Somerville, G. A.; Peters, G.; Sinha, B.; Harraghy, N.; Proctor, R. A.; Herrmann, M. *J Bacteriol* **2005**, *187*, 4488.
- (26) Valihrach, L.; Demnerova, K. *J Microbiol Methods* **2012**, *90*, 214.
- (27) Kovacs, M.; Halfmann, A.; Fedtke, I.; Heintz, M.; Peschel, A.; Vollmer, W.; Hakenbeck, R.; Bruckner, R. *J Bacteriol* **2006**, *188*, 5797.
- (28) Schulthess, B.; Bloes, D. A.; Francois, P.; Girard, M.; Schrenzel, J.; Bischoff, M.; Berger-Bachi, B. *J Bacteriol* **2011**, *193*, 4954.
- (29) Halouska, S.; Zhang, B.; Gaupp, R.; Lei, S.; Snell, E.; Fenton, R. J.; Barletta, R. G.; Somerville, G. A.; Powers, R. *Revisiting Protocols for the NMR Analysis of Bacterial Metabolomes*, 2013.
- (30) Zhang, B.; Halouska, S.; Schiaffo, C. E.; Sadykov, M. R.; Somerville, G. A.; Powers, R. *J Proteome Res* **2011**, *10*, 3743.
- (31) Bischoff, M.; Dunman, P.; Kormanec, J.; Macapagal, D.; Murphy, E.; Mounts, W.; Berger-Bachi, B.; Projan, S. *J Bacteriol* **2004**, *186*, 4085.
- (32) Deora, R.; Misra, T. K. *J Biol Chem* **1996**, *271*, 21828.
- (33) Rao, L.; Karls, R. K.; Betley, M. J. *J Bacteriol* **1995**, *177*, 2609.
- (34) Maddocks, S. E.; Oyston, P. C. *Microbiology* **2008**, *154*, 3609.
- (35) Somerville, G. A.; Proctor, R. A. In *Staphylococci in Human Disease*; Wiley-Blackwell: 2009, p 1.
- (36) Strasters, K. C.; Winkler, K. C. *J Gen Microbiol* **1963**, *33*, 213.

- (37) Seidl, K.; Muller, S.; Francois, P.; Kriebitzsch, C.; Schrenzel, J.; Engelmann, S.; Bischoff, M.; Berger-Bachi, B. *BMC Microbiol* **2009**, *9*, 95.
- (38) Somerville, G. A.; Chaussee, M. S.; Morgan, C. I.; Fitzgerald, J. R.; Dorward, D. W.; Reitzer, L. J.; Musser, J. M. *Infect Immun* **2002**, *70*, 6373.
- (39) Coulter, S. N.; Schwan, W. R.; Ng, E. Y.; Langhorne, M. H.; Ritchie, H. D.; Westbrook-Wadman, S.; Hufnagle, W. O.; Folger, K. R.; Bayer, A. S.; Stover, C. K. *Mol Microbiol* **1998**, *30*, 393.
- (40) Mei, J. M.; Nourbakhsh, F.; Ford, C. W.; Holden, D. W. *Mol Microbiol* **1997**, *26*, 399.
- (41) Begun, J.; Sifri, C. D.; Goldman, S.; Calderwood, S. B.; Ausubel, F. M. *Infect Immun* **2005**, *73*, 872.
- (42) Bae, T.; Banger, A. K.; Wallace, A.; Glass, E. M.; Aslund, F.; Schneewind, O.; Missiakas, D. M. *Proc Natl Acad Sci U S A* **2004**, *101*, 12312.
- (43) Chaffin, D. O.; Taylor, D.; Skerrett, S. J.; Rubens, C. E. *PLoS One* **2012**, *7*, e41329.
- (44) Sadykov, M. R.; Mattes, T. A.; Luong, T. T.; Zhu, Y.; Day, S. R.; Sifri, C. D.; Lee, C. Y.; Somerville, G. A. *J Bacteriol* **2010**, *192*, 1459.
- (45) Zhu, Y.; Xiong, Y. Q.; Sadykov, M. R.; Fey, P. D.; Lei, M. G.; Lee, C. Y.; Bayer, A. S.; Somerville, G. A. *Infect Immun* **2009**, *77*, 4256.

CHAPTER 9

OVEREXPRESSION AND PURIFICATION OF ISOTOPICALLY LABELED C-TERMINAL DOMAIN OF RPIRA FOR NMR STUDIES

9.1 Introduction

Staphylococcus aureus is a life-threatening human pathogen associated with biofilm formation on implantable medical devices.¹ Biosynthesis of virulence factors that can determine the pathogenicity of staphylococci is found to be regulated by tricarboxylic acid (TCA) cycle activity.² An interesting observation in Chapter 6 is that the TCA cycle, as a component of central metabolism, can interdependently influence the carbon flow of glycolysis and the pentose phosphate pathway (PPP).³ The fact that intracellular ribose concentration increases during TCA cycle suppression indicates the presence of a pentose phosphate pathway (PPP)-responsive regulator that can partially mediate TCA cycle-dependent regulatory effects.⁴ This led us to hypothesize that the ribose phosphate isomerase (Rpi) protein is a potential regulator of PPP and TCA cycle.⁵⁻⁷

Ribose can be phosphorylated and uptaken from the growth medium using an Rbs transport system or generated from nucleotide degradation.⁸ Ribose 5-phosphate is mainly involved in the non-oxidative branch of the pentose phosphate pathway (PPP): ribose 5-phosphate, ribulose 5-phosphate and xylulose 5-phosphate are interconverted in reactions catalyzed by ribose phosphate isomerase and ribulose-5-phosphate 3-epimerase (Figure 9.1).⁹ The family of Rpi naturally occurs in two distinct classes, ribose phosphate isomerase A (RpiA; EC 5.3.1.6) and ribose phosphate isomerase B (RpiB; EC 5.3.1.6),

and structures of both has been resolved using X-ray crystallography in several organisms.¹⁰⁻¹² In *Escherichia coli* (*E. coli*), RpiA is a homodimeric enzyme with an $\alpha/\beta/(\alpha/\beta)/\beta/\alpha$ fold where arabinose 5-phosphate was identified as an inhibitor and the complex structure was previously determined.¹¹ RpiB is unrelated to RpiA in terms of protein sequence and structure. RpiB exhibits a Rossmann-type $\alpha\beta\alpha$ -sandwich fold and also forms a dimer in *E. coli* and *Coccidioides immitis* (*C. immitis*).^{10,13} Besides the same catalytic function as RpiA, a secondary activity of RpiB is unknown.

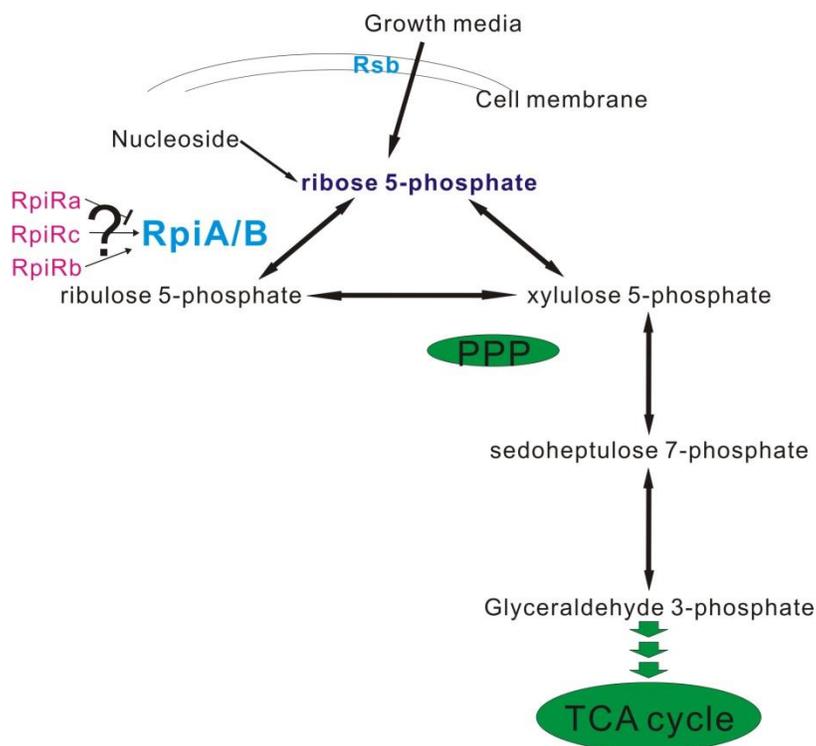


Figure 9.1 RpiR family of proteins role in the central metabolism.

Members of the RpiR protein family often act as transcriptional regulators of sugar catabolism, and RpiR homologues have been identified as repressors and activators in both Gram-negative and Gram-positive bacteria, including *E. coli*, *Pseudomonas putida*, and *Bacillus subtilis*.^{9,14,15} As sugar-responsive regulators, members of the RpiR family of proteins have N-terminal helix-turn-helix DNA binding motifs and C-terminal sugar isomerase binding domain.¹⁶ However, detailed structural information has not been reported. There are three known members of the RpiR family: RpiRa, -b, and -c. RpiRb and -c positively influence the activity of RpiA while RpiRa works as a downstream antagonist of RpiA by binding to a certain dinucleotide (Figure 9.1).⁴ It has also been reported that the inactivation of RpiRc altered the temporal transcription of RNAIII, the effector molecule of the *agr* quorum-sensing system.⁴ Thus, central metabolism is potentially linked with virulence determinant synthesis through RpiR and PPP.⁴ Consequently, our initial research goal is the structural determination and functional annotation of the *S. aureus* RpiRa protein (*S. aureus subsp. aureus* Mu50, UniProt ID: Q99WQ9; Gene Symbol SAV0317; Gene ID: 1120275) by NMR.

NMR spectroscopy is one of the cornerstone techniques for structural studies of biological macromolecules. It is the only current technique that can obtain atomic level information about both the structure and dynamics of macromolecules such as proteins, nucleic acids and their complexes in a solution. The limitations of NMR spectroscopy result from the low inherent sensitivity and from the high complexity of NMR spectra. As a specific example, a minimal of 10-20 mg/L expression yield is practically required for sufficient sensitivity in three and four dimensional NMR experiments. NMR also has a

molecular-weight limitation and a routine strategy is to focus the NMR analysis to the smallest possible subdomain. Thus, only the ligand binding C-terminal domain (CTD) of RpiRa will be studied by multi-dimensional NMR. Two-, three-, and four-dimensional NMR experiments are required to disperse the hundreds of NMR resonances from RpiRa CTD to avoid spectral overlap and simplify the analysis. Typically, the additional NMR dimensions correspond to ^{13}C and ^{15}N chemical shifts that are correlated with ^1H chemical shifts through large single-bond coupling constants. Heteronuclei like ^{15}N , ^{13}C and ^2H can be incorporated into proteins by uniform or selective isotopic labelling. As a result, the NMR spectra are drastically simplified and the sensitivity of the nitrogen and carbon signals is largely increased. In summary, obtaining a well folded protein with a minimal molecular-weight that can be overexpressed and purified at a high concentration and with isotopic labeling is the first step of using NMR to study a protein structure. Several challenging aspects of this process that include overproduction, purification, storage and isotopic labeling are addressed in this chapter.

The optimization of protein expression in a heterogenous system is one of the major challenges of structural biology. Many recombinant proteins are expressed in the prokaryotic system of *E. coli* due to its simplicity and high efficiency, but not all proteins are expressed equally well in *E. coli*.¹⁷ Thus, it is critical to optimize the experimental conditions in order to maximize the protein yield. Many different strategies have been developed to meet the needs of a variety of proteins, such as *E. coli* strains supplemented with rare codons,¹⁸ strains with disabled thioredoxin reductase and glutathione reductase,¹⁹ plasmids with a tunable promoter,²⁰ transport or periplasmic spaces,²¹ and

co-expression with other proteins.²² Despite all these efforts that have successfully addressed known mechanisms of low yield, “trial and error” is still a large component of most protein expression efforts.

Another common challenge of a protein structural study is the requirement for a highly purified protein sample. In general, a purity level of 90%-95% of the target protein is needed. A fusion tag is often used to simplify the purification process.²³ An affinity tag is a recombinant hybrid containing a polypeptide fusion partner which facilitates the purification of the target protein.²⁴ There are many different types of tags and the choice of tag depends on the property of the protein: poly-Arg, FLAG, poly-His, c-myc, S and Strep II-tag.²⁵ A poly-His tag utilizes an immobilized metal affinity chromatography (IMAC) to purify a recombinant protein containing hexa-histidine residues as a short affinity-tag. A His-tag has been carried out successfully using a number of expression systems including bacteria, yeast, and mammalian cells. It is commercially incorporated in the cloning vectors such as pET-28a-c (+), which we used to express RpiRa CTD.

A purified protein has to be stored in a buffer at an extremely high concentration of 1-2 mM for NMR studies. Therefore, another challenge of NMR structural biology is the identification of an optimal buffer condition that can stabilize a properly folded protein at a high concentration. The buffer should exhibit little or no change in pH with temperature, show insignificant penetration through biological membranes, and have maximum buffer capacity at the pH where the protein exhibits optimal stability. The isoelectric point, hydrophobicity, intermolecular interactions and many other protein

properties and NMR requirements significantly impacts and limits buffer choices. There is no reliable approach to predict an optimal buffer for a given protein. Some of the most commonly used buffer systems include Tris-HCl, sodium/potassium phosphate saline, MES, and acetate.²⁶ There are other factors to consider that include salt concentration, pH, reducing reagents, etc. In short, buffer composition and physical properties can perturb protein conformational stability because of a complex interplay between various effects rather than through a single mechanism. The goal is to identify an optimal buffer that is beneficial to both the protein and the NMR analysis.

One of the most important techniques for NMR structural studies is isotopic labeling. The introduction of uniform ^{15}N and ^{13}C labeling into a protein provides NMR signals that can be sequentially assigned by “walking through” correlated peaks between neighboring residues. A two dimensional (2D) ^1H , ^{15}N heteronuclear single quantum coherence (HSQC) spectrum is used as a template in which each peak corresponds to the amide H-N bond for each amino acid residue in the protein sequence.²⁷ Thus, the 2D ^1H , ^{15}N HSQC spectrum provides the “root” ^1H and ^{15}N chemical shifts for each amino acid in the protein. Each root is then connected to additional ^1H and ^{13}C chemical shifts through a series of three dimensional NMR spectra. This establishes a complete set of peaks per residue (spin system) based on the 2D ^1H , ^{15}N HSQC spectrum.²⁷ Sequential assignments are then made by identifying spin systems (residues) that have two or more chemical shifts in common. Therefore, the structural analysis of proteins using NMR starts with a 2D ^1H , ^{15}N HSQC spectrum. A high quality, well dispersed 2D ^1H , ^{15}N HSQC spectrum indicates a properly folded protein. It also identifies the likelihood that

the NMR assignments can be completed. The availability of backbone sequential assignments and a high-quality 2D ^1H , ^{15}N HSQC spectrum are also important for an NMR ligand-affinity screen (FAST-NMR).²⁸ So, a 2D ^1H , ^{15}N HSQC spectrum is the primary means of evaluating the preparation of a protein sample for NMR.

Deuterium (^2H) labeling can simplify an NMR spectrum and sharpen peaks by removing pathways that contribute to transverse relaxations. This is critical for studying proteins with a high molecular weight because larger molecules tend to have broader peaks due to their relatively slower tumbling and complex dipole-dipole relaxations. Incorporating ^2H labeling into a protein sample is difficult and more challenging compared to ^{15}N or ^{13}C isotope labeling protocols. ^2H doubles the atomic weight that leads to significant isotope effects and drastic changes in enzymatic activity. As a result, the viability of *E. coli* is negatively impacted by growing cells in D_2O .²⁹ If D_2O is the sole deuterium source, 0–86% of the chemically nonexchangeable hydrogen atoms can be deuterated. Higher levels of “perdeuteration” require using deuterated glucose because cell growth in higher percentages of D_2O is not practical. In addition to these uniform labeling strategies, selective labeling methods have also been applied for more challenging systems to further simplify the NMR spectrum.³⁰ These isotopic approaches complement NMR techniques to explore protein structures. For example, transverse relaxation optimized spectroscopy (TROSY) type experiments can be used with a high-field magnet NMR to increase the sensitivity of an NMR spectrum by selecting the component, from a decoupled spectrum, in which the relaxation mechanisms have almost cancelled leading to a single, sharp peak.³¹ TROSY is often

combined with the ^2H labeling.^{32,33} The efforts to optimize the overexpression and purification of ^2H , ^{13}C , ^{15}N labeled RpiRa CTD, and the use of selective labeling schemes along with the application of TROSY-based NMR experiments for the backbone sequential assignments of RpiRa CTD are discussed in this chapter.

9.2 Materials and Methods

9.2.1 Materials

For the RpiRa CTD₇₉₋₂₆₆ NMR analysis, purified and uniformly ^{15}N labeled, uniformly ^{13}C , ^{15}N double labeled, and uniformly ^2H , ^{13}C , ^{15}N , triple labeled proteins were expressed in a pET-28a-c(+) cloning plasmid with an N-terminal His-Tag sequence in the *E. coli* BL21 (DE3) strain. N-terminal modification of RpiRa CTD₇₉₋₂₆₆ included the addition of the following amino acid sequence: MGSSHHHHHSSG LVPRGSHMAS.

Deuterium oxide (99.9% D), ^{15}N - NH_4Cl (98% ^{15}N), $^{13}\text{C}_6$ -glucose (99% D), imidazole, NH_4OAc , yeast extract Hy-Yest® 444, kanamycin sulfate salt, DL-dithiothreitol, sodium azide (99.5%), and glycerol were obtained from Sigma-Aldrich (Milwaukee, WI). The 3-(trimethylsilyl) propionic-2, 2, 3, 3-d₄ acid sodium salt (98% D) was purchased from Cambridge Isotope (Andover, MA). The potassium phosphate dibasic salt (anhydrous, 99.1% pure) and sodium chloride (ACS reagent) was purchased from Mallinckrodt (Phillipsburg, NJ). The Bacto™ Agar and Bacto™ Typtone were purchased from BD (Franklin Lakes, NJ). The prestained Protein Marker, Broad Range

(7-175 kDa) was purchased from New England BioLabs (Ipswich, MA). The tris (2-carboxyethyl) phosphine hydrochloride (TCEP) was purchased from Biosynth (Itasca, IL). The isopropyl β -D-thiogalactoside (IPTG) was purchased from GOLDBIO®COM (St. Louis, MO). All the recipes for the cell culture media and buffers are listed in the Appendix B.

9.2.2 Plasmid construction, *E. coli*. strains and glycerol starter

The three plasmids with three different RpiRa CTD amino acid sequences and the corresponding *E. coli* BL21 (DE3) strains were prepared by Dr. Marat Sadykov, Dr. Nagender Ledala and Mckenzie Steger in Dr. Somerville's lab.

Selected *E. coli* BL21 (DE3) colonies from fresh agar plate were inoculated into 3 mL of Luria-Bertani (LB) media containing 30 μ g/mL kanamycin. The cells were incubated at 37°C overnight with shaking at 220 rpm in a GYROMAX™ incubator shaker (Concord, CA). 0.5 mL of glycerol was added to 1 mL of the initial culture, and the mixture as the glycerol stock was stored at -80°C.

Same protocol was used to generate D₂O adapted cultures. A series of minimal media containing 20%, 40%, 60% and 80% of D₂O were used sequentially to adapt the cell culture to 80% D₂O concentration. No antibiotics were added to the minimal media. DNA sequencing was performed after this culturing to validate the sequence.

9.2.3 Protein expression in M9 medium and high cell density medium

Selected colonies from fresh agar plate were inoculated into 3 mL of LB media containing 30 µg/mL kanamycin. The cells were incubated at 37°C overnight with shaking at 220 rpm. 1 mL initial culture was diluted into 100 mL M9 minimal media containing 30 µg/mL kanamycin. The culture was grown at 37°C and 220 rpm until an optical density at 600 nm (O.D.₆₀₀) reached 0.6-0.8 unit. Freshly prepared stock solution of IPTG was added to make a final concentration of 0.8 mM. The cells were grown for another 3 hours. The cells were harvested by centrifuging at 5000 X g for 20 minutes at 4 °C in a Sorvall RC-5B Refrigerated Superspeed centrifuge (Du Pont Instruments, Wilmington, DE) and the pellets were collected for protein extraction.

Same incubation and centrifugation settings were used for high-cell density medium-based protein expression. Selected colonies from fresh agar plate were inoculated into 3 mL of LB media containing 30 µg/mL kanamycin. The cells were incubated at 37°C overnight with shaking at 220 rpm. 1 mL initial culture was diluted into 100 mL LB medium containing 30 µg/mL kanamycin. The culture was grown at 37°C and 220 rpm until an O.D.₆₀₀ reached 2.5-3.0 units. The cells were harvested by centrifuging at 4000 X g for 8 minutes. The cell pellets were suspended in high cell density medium and inoculated for another 0.5-1 hour until an O.D.₆₀₀ was increased by 1 unit. IPTG was added with a final concentration of 1 mM and the cells were grown for another 4 hours. The cells were harvested by centrifuging at 5000 X g for 20 minutes at 4 °C and the pellets were collected for protein extraction.

9.2.4 Protein extraction and purification

The cells were re-suspended in the protein extraction buffer (see recipe in the Appendix B) of 15-25 mL. Sonication was performed to lyse the cells using a Fisher Scientific sonicator probe, Sonic Dismembrator Model 100 (Waltham, MA). A burst cycle of 15 seconds on and 45 second off were chosen with 60% power. Nine burst cycles were completed before the lysates were spun down at 7,000 X g for 45 min at 4°C. The sample was kept on ice during this process.

The extraction was followed by purification. A gravity-flow IMAC column packed with HisPur Cobalt beads (Thermo scientific, Rockford, IL) was equilibrated with 25 mL wash buffer consisting of phosphate buffer saline pH 7.2 with 10 mM imidazole (to prevent nonspecific binding of endogenous proteins that have histidine clusters). The cell lysate was applied to the column and allowed to bind for 2-3 h during which the lysate solution was thoroughly mixed with the beads. The column was repositioned vertically and drained to collect the flow-through. The beads were then washed with two wash buffers: first with ~200 mL of 10 mM imidazole and second with ~200 mL of 50 mM imidazole. The bound RpiRa CTD₇₉₋₂₆₆ was eluted by applying ~300 mL elution buffer with 150 mM imidazole. The flow rate was 1 mL/min and 1 mL fractions were collected. The presence of RpiRa CTD₇₉₋₂₆₆ in the eluted fractions was assayed by Ultraviolet-visible (UV-Vis) spectroscopy at a wavelength of 280 nm. All purification steps were performed at 4 °C.

9.2.5 Imidazole removal, protein concentration and NMR sample preparation

All eluted fractions were collected until UV-Vis absorptive spectral readings reached baseline. These fractions were concentrated to 10 mL using 15 mL Amicon ultrafiltration cells with a 10 kDa cutoff membrane (EMD Millipore, Billerica, MA). The filtration was performed by centrifugation using a Thermo scientific CL10 bench-top centrifuge (Rockford, IL) and a swinging bucket rotor at 4000 x g and 25 °C. A 10 mL Float-A-Lyzer G2 with 8-10 kDa cutoff membrane from Spectrum® Laboratories (Rancho Dominguez, CA) was used for dialysis. A total of 2 L NMR buffer was prepared and four rounds of dialysis with 200, 200, 600, 1000 mL volume of the buffer were performed. The protein was concentrated to 8-10 mL and stored at room temperature. The RpiRa CTD₇₉₋₂₆₆ was concentrated to a final volume of 500 µL using a 0.5 mL Amicon ultrafiltration cells. 50 µL of a D₂O buffer was added to the concentrated RpiRa CTD₇₉₋₂₆₆ protein sample to yield an NMR sample consisting of 1.2 mM RpiRa CTD₇₉₋₂₆₆ in a 91% H₂O/9% D₂O solution containing 100 mM NaCl, 25 mM K₂HPO₄/KH₂PO₄ at pH 4.5 (uncorrected). An Eppendorf 5415D Micro Centrifuge (New York, NY) was used at 4000 X g at 25°C to concentrate the sample. The sample was kept in 5 mm NMR tubes (Norell ST500-7, Norell, Inc., Landisville, NJ).

9.2.6 Analytical Apparatus and data analysis

For the RpiRa CTD₇₉₋₂₆₆ NMR analysis, one dimensional (1D) ¹H NMR experiments were collected at 298 K on a three channel 500 MHz Bruker Avance DRX

spectrometer equipped with a triple resonance, Z-axis gradient 5 mm Cryoprobe. A total of 8192 data points were collected with a total of 16 dummy scans and 128 acquisition scans and a sweep width of 5482.5 Hz. 2D ^1H , ^{15}N HSQC NMR experiments used were collected at 298 K on the same instrument. A total of 1024 data points were collected in the ^1H dimension and 128 data points were collected in the ^{15}N dimension. The spectrum was collected with a total of 256 dummy scans and 64 acquisition scans and a sweep width of 7002.8 Hz in the ^1H dimension and 2027.3 Hz in the ^{15}N dimension. TROSY-2D ^1H , ^{15}N HSQC was collected at 298 K on a three channel 700 MHz Bruker Avance III-HD spectrometer equipped with a triple resonance, Z-axis gradient 5 mm Cryoprobe. A total of 2048 data points were collected in the ^1H dimension and 128 data points were collected in the ^{15}N dimension. The spectrum was collected with a total of 128 dummy scans and 32 acquisition scans and a sweep width of 9803.9 Hz in the ^1H dimension and 2838.4 Hz in the ^{15}N dimension. 1D ^1H NMR spectra were processed in the ACD/1D NMR manager version 12.0 (Advanced Chemistry Development, *Inc*). 2D ^1H , ^{15}N HSQC spectra were processed using the NMRPipe software package.³⁴

UV-Vis spectroscopy was used for protein concentration determination and optical density of cells using a PerkinElmer Lambda Bio+ spectrometer (Waltham, MA). Circular dichroism (CD) absorption spectra were recorded on a JASCO J-815 spectropolarimeter (Easton, MD). A wavelength range of 200–240 nm in far-UV region was used for probing the secondary structure of RpiRa CTD₇₉₋₂₆₆. Far-UV CD spectra were collected with 1 mm path length cuvette on an RpiRa CTD₇₉₋₂₆₆ sample with an estimated concentration of 50 μM in 50 mM potassium phosphate buffer, pH 7.4. All CD

spectra were acquired at a scan speed of 10 nm/ min. The spectrum was plotted in Excel (Microsoft, Redmond, WA).

SDS–polyacrylamide gel electrophoresis of the proteins in all fractions was performed using a 4%-12% PAGE with a Bio-Rad Mini-Protean® 3 Multi-Casting Chamber (Hercules, CA). Protein bands were visualized by Coomassie brilliant blue staining and photographed.

9.3 Results

9.3.1 Determination of CTD amino acid region

```

      10      20      30      40      50
MKFENRVQRY QHLFTKTDKR IVNYIRQNGY SDAFSTINSL AHAIGTSPAT
      60      70      80      90     100
MTRFSHKLDY ENFQDLKFNI QQEMTETVIE NSPIIQRIHK YHQQIIQQTG
      110     120     130     140     150
EFIDNDIIQT FIDKLQSSRH ILFAGLGSSG LSATEFYIRM IRMGLKGNVT
      160     170     180     190     200
TDSHLMKISA SLLSHSDMFI AMSNSGNTSE LISAAEVAKS HGAYVVAITN
      210     220     230     240     250
FEGSKLTDCA DLVLLTTDQS RNTDHFINT QIATLFLIDI VSYHLENTN
      260
LSQTYQHTKS IILDNK

```

Scheme 9.1 RpiRa sequence.

S. aureus RpiRa has a single polypeptide chain containing 266 amino acid residues. The complete protein primary sequence is shown in Scheme 1. A collection of database including DomPred,³⁵ InterProScan³⁶ and DOMAC³⁷ were performed and two domains were predicted: an N-terminal domain (NTD) of helix-turn-helix-like domain and a C-terminal domain (CTD) of sugar isomerase-like domain. The predicted boundaries of the CTD were shown below (Table 9.1). The starting position of CTD was predicted with a range of 78-113. A closer look at the secondary structure prediction using PsiPred,^{35,38} SWISS-MODEL³⁹ and PredictProtein⁴⁰ suggested that the boundary of CTD and NTD should be in loop regions of 75-82 or 112-123. So a series of constructs were proposed and three were validated experimentally. It turned out that both PsiPred provided the closest prediction that residue I79 is the starting amino acid of an α -helix in CTD and therefore the secondary prediction based on the same database was used for secondary structure prediction.⁴¹

Table 9.1 Summary of boundary prediction of RpiR CTD.

	Beginning	Ending	E-value
BLAST	108	240	6.11e-37
FASTA	112	238	5.1e-105
InterproScan	111	251	6.8e-13
DomPred	113	NA	NA
PsiPred	79	266	2e-07
SWISS-MODEL	78	264	NA

Three forms of RpiRa were constructed to experimentally determine the CTD boundaries by residue sequence numbers: 81-259, 79-249 and 79-266 and the corresponding proteins were termed as RpiRa CTD₈₁₋₂₅₉, CTD₇₉₋₂₄₉, and CTD₇₉₋₂₆₆. These

constructs were amplified by polymerase chain reaction (PCR) and cloned separately into plasmid pET-28a (+). The CTD₇₉₋₂₆₆ strain was selected based on its relatively high protein expression level. The CTD₇₉₋₂₆₆ strain has the longest sequence of the three constructs and contains the seven predicted α -helices, increasing the likelihood of obtaining correctly folded protein. There are 188 amino acids in the predicted RpiRa CTD₇₉₋₂₆₆ with a molecular weight of 21009.7 kDa. RpiRa CTD₇₉₋₂₆₆ is also predicted to have an extension coefficient of 8940 M⁻¹ cm⁻¹ at 280 nm, assuming all Cys residues are reduced. The theoretical isoelectric point (pI) is 5.95. These parameters were calculated using ProtParam.⁴² The experimental CD spectrum of RpiRa CTD₇₉₋₂₆₆ predicted that RpiRa CTD₇₉₋₂₆₆ is 50.4% α -helical and 15.5% β -stranded (Figure 9.2b). This is also comparable with the RpiRa homology modeling results (α -helix%: 51.1; β strand%: 13.3) using PsiPred (Figure 9.2a).³⁸

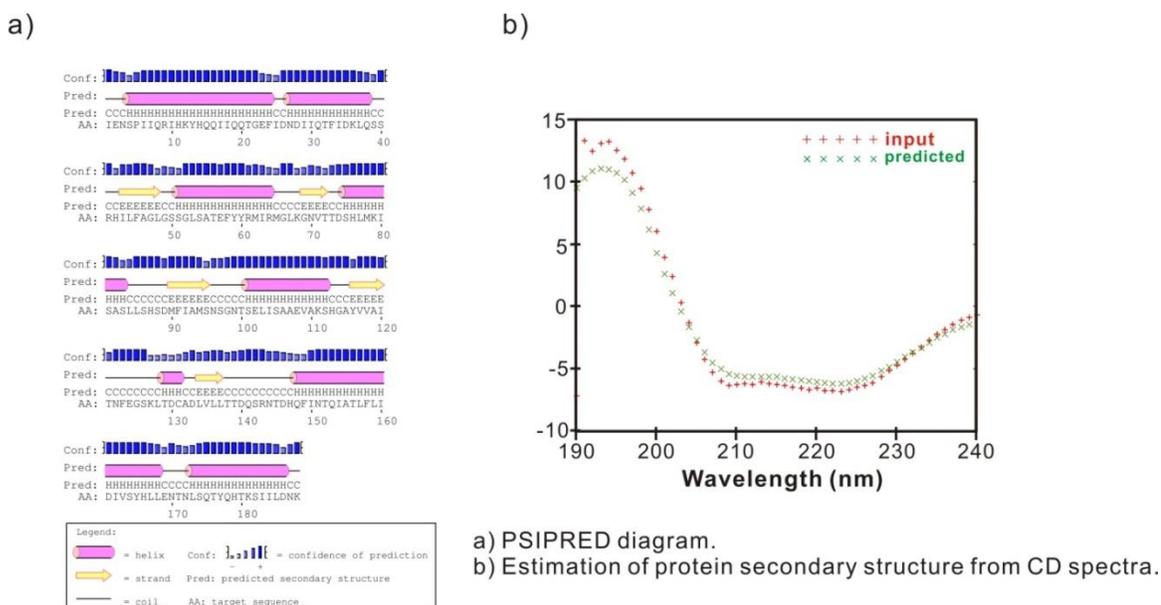


Figure 9.2 a) Secondary structure prediction using PSIPRED

(<http://bioinf.cs.ucl.ac.uk/psipred/>). Top matched structure is *Vibrio vulnificus* NanR protein (PDB ID: 4IVN) with a percentage identity of 29%. b) CD spectrum prediction using K2D2, a web server to estimate α helix and β strand content of a protein from its CD spectrum. (<http://k2d2.ogic.ca/>). Red line represented the query spectrum, and green line represented the target spectrum. The maximum total α and β error of the prediction was 0.32.

9.3.2 Protein expression

The optimization of protein expression was focused on optimizing cell growth conditions. M9 and LB media only yielded an O.D.₆₀₀ of 1.5-1.8 units. Since the protein expression level was also low; the limited amount of cells became a bottleneck for providing enough proteins in a standard 1 L culture. Therefore, a high cell density medium was used (Appendix B).⁴³ Cell mass was increased to four folds compared to

normal M9 media. The procedure is outlined in Figure 9.3a. The growth curves (induced vs. non-induced) based on medium switch are shown in Figure 9.3b. A large production of cells was achieved without wasting unnecessary nutrients. The media switch was also beneficial for the introduction of ^{13}C glucose and/ or ^{15}N ammonium chloride for isotope labeling. Bacterial cells in their mid-exponential phase were induced by IPTG and grown to a higher O.D.₆₀₀ of 5-6 units. A comparison between the protein expressions indicates that a high cell density growth provided a higher protein yield (Figure 9.3c).

The protein was also expressed under anaerobic growth and aerobic growth conditions. A volume ratio of flask to medium at 2:1 was used to generate an anaerobic condition, while the ratio at 8:1 was used for aerobic growth. As shown in Figure 9.2c, the expression level was much better under aerobic growth using the high cell density growth. These results also agree with previous reports that high cell density media for *E. coli* are generally designed for aerobic cultivations.⁴⁴ A room temperature overnight expression and a prolonged time of expression were also tested. Neither approach was shown to be beneficial (Figure 9.4) since no increase in expression levels were seen after 4 hours. Different IPTG concentrations of 0.4, 0.6, 0.8, 1.0 and 1.2 mM were also investigated, but no significant changes were observed. An IPTG concentration of 1.0 mM was used for all subsequent protein sample preparations.

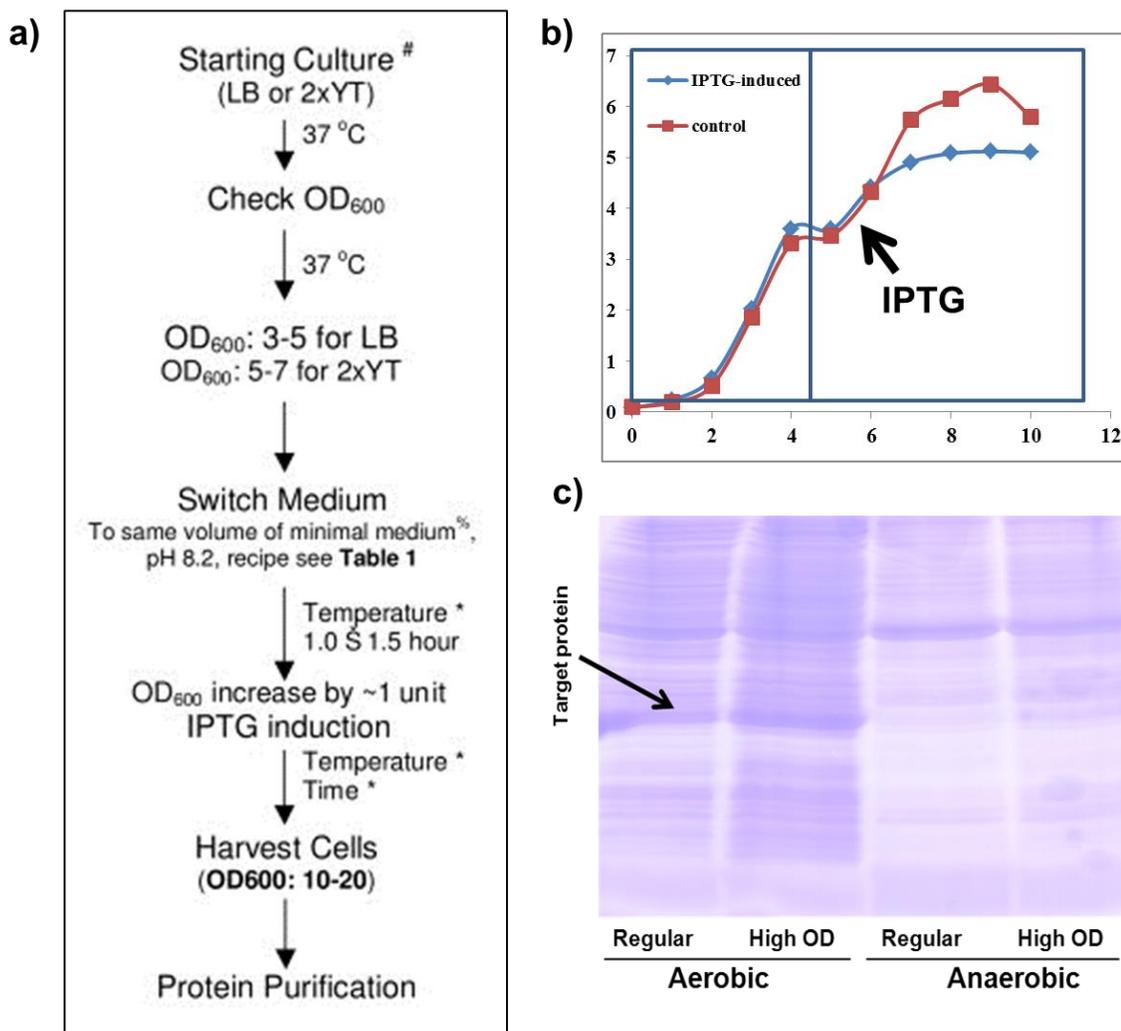


Figure 9.3 The overexpression of RpiRa CTD₇₉₋₂₆₆: a) flowchart illustrating the protocols for the high-cell-density expression system; b) *E. coli* growth curve using the high-cell-density expression system; c) comparison of RpiRa CTD₇₉₋₂₆₆ protein expression levels under different growth conditions.

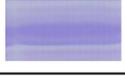
Time	pH	RpiRA
1	6.75	
2	6.61	
3	6.33	
4	6.4	
5	6.31	
6	6.26	
15	5.92	

Figure 9.4 Impact of pH changes on RpiRa CTD₇₉₋₂₆₆ protein expression levels using aerobic growth with high cell density media at 25°C.

9.3.3 Protein extraction

Cell lysis was accomplished using sonication. The cell sample was immersed in an ice bath during sonication to prevent any overheating effects. 20-25 mL extraction buffer was used and the choice of buffer was mainly dependent on the compatibility of the following purification procedures. Therefore 50 mM potassium dihydrogen phosphate/ dipotassium hydrogen phosphate buffer (KH₂PO₄/K₂HPO₄) was chosen with a pH of 7.4. It is advisable to measure the pH during the cell lysis to maintain a neutral or mild basic conditions.⁴⁵ 5% glycerol was also proved to be useful to prevent damage to the protein during the sonication. A burst cycle of 15 seconds on and 45 second off was

chosen and sonication powers from 40% to 80% were tested to maximize the efficiency of cell lysing. A 60% sonication amplitude was chosen due to its high lysing efficiency without generating the appearance of a foam for a 20-25 mL cell suspension. The presence of foam indicates denatured or destroyed proteins. As shown in Figure 9.5, extending the number of burst cycles increased the amount of RpiRa CTD₇₉₋₂₆₆ being released from the cells. Nine burst cycles were chosen because it corresponded to the largest increase in RpiRa CTD₇₉₋₂₆₆ with the smallest number of burst cycles.

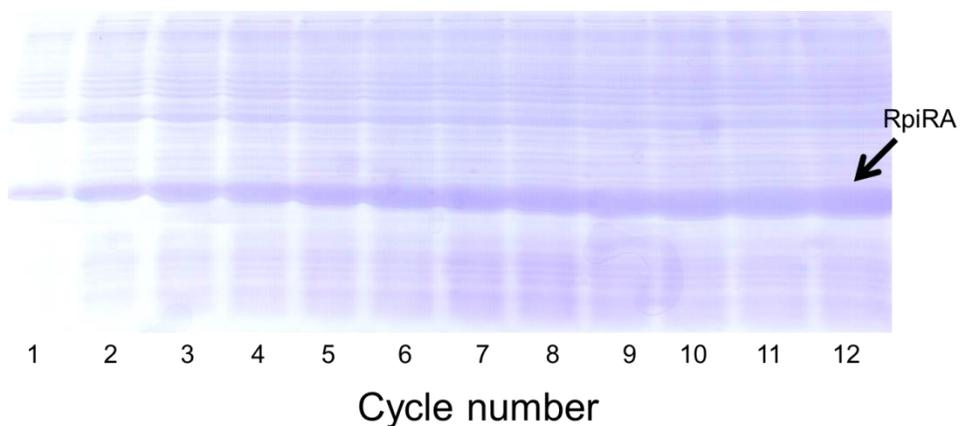


Figure 9.5 The impact of the number of burst cycles on the efficiency of extracting of RpiRa CTD₇₉₋₂₆₆.

9.3.4 Protein purification

HisPur cobalt IMAC resin can bind more than 10 mg proteins per mL of resin and can maximize protein purity compared to nickel resins. The exact protocol was followed.

The only variable in regards to the standard protocol was the length of the binding time – the time the cell lysate is simply incubated with the resin. Binding times of 2, 4 and 8 hours were tested, but 2 hours was chosen since extending the binding time did not improve either the relative protein purification or the amount of protein recovered. Figure 9.5 showed the purified protein band in a denatured SDS-PAGE. The flowthrough sample in Lane 2 indicated almost all the target protein was captured and the target protein band in Lane 7 showed that the purity of the protein and the correct molecular weight in the non-native SDS-PAGE.

The protein band in the elution buffer as shown in Figure 9.6a was rather pure, but an extremely concentrated and overloaded sample (Figure 9.6b) verified the purity of RpiRa CTD₇₉₋₂₆₆. After weeks of room-temperature storage, a small amount of proteolytic degradation was observed. Therefore, it is advisable to include broad-spectrum protease inhibitors in the extraction buffer and the NMR sample. Figure 9.6b demonstrates that after IPTG induction, not only was RpiRa CTD₇₉₋₂₆₆ overexpressed, but the levels of other protein were decreased. So, it appears that under carbon limitation, the induction of a protein overexpression also suppress other protein expression. All of these results are consistent with our protocol being successful in inducing the expression of soluble RpiRa CTD₇₉₋₂₆₆ exclusively and efficiently.

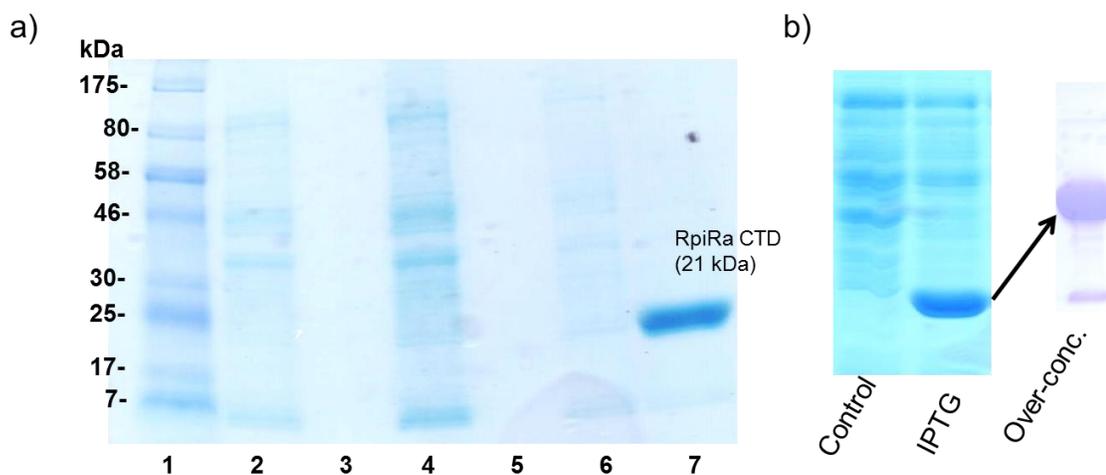


Figure 9.6 a) The Co-affinity chromatography purification of RpiRa. Lane 1 is the protein ladder, Lane 2 is the flow-through of non-binding proteins, Lane 4 and 6 are two concentrated wash buffers containing non-specific bound proteins, Lane 7 is the elution buffer containing the target protein. b) The cell lysate, before purification, that compares expression and non-expression conditions. An over-loaded gel of RpiRa CTD₇₉₋₂₆₆ was also performed to verify the purity of the protein.

9.3.5 Protein buffer conditions

The pH was chosen to be from 4.5-5.5 so that the pH was at least one unit lower from the theoretical pI value of 5.95. A lower pH also benefits the quality of the 2D ¹H, ¹⁵N-HSQC spectrum by decreasing backbone amide exchange rates.²⁷ A series of standard NMR protein buffer were tested that included: Tris buffer, MES buffer, NH₄OAc, acetate and potassium phosphate buffers (Appendix B). NH₄OAc and phosphate buffers both yielded higher quality NMR spectra. However, a phosphate buffer was chosen for the final NMR sample conditions because of slightly better protein

stabilization at higher temperatures. NaCl concentrations of 100 mM, 200, 300, 400, 500 mM were also tested. The final NMR sample contained 100 mM of NaCl because higher ion strengths (>300 mM NaCl) was shown to denature the protein or decrease the performance of the cryoprobe without any added benefit to protein stability. Additives such as imidazole and glycerol may also improve protein stability, but they can also negatively impact the quality of the NMR spectrum.⁴⁶ No obvious benefit was observed for including any additive to the RpiRa CTD₇₉₋₂₆₆ NMR sample (Figure 9.7).

Dithiothreitol (DTT) was also tested over a range of concentrations: 1, 5 and 10 mM. 10 mM DTT was proved to be necessary to the long-term stability of the RpiRa CTD₇₉₋₂₆₆ NMR sample. An alternative reducing reagent, TECP-HCl, was shown to similarly stabilize the RpiRa CTD₇₉₋₂₆₆ NMR sample at a concentration of 2 mM. A typical 1D ¹H NMR spectrum for RpiRa CTD₇₉₋₂₆₆ in the optimized NMR buffer is shown in Figure 9.8.

A flash frozen and a slow frozen approach were tested for possible approaches to achieve a long-term storage of RpiRa CTD₇₉₋₂₆₆. Unfortunately, RpiRa CTD₇₉₋₂₆₆ was shown to be denatured after thawing from either these long-term storage techniques. A storage temperature of -20°C or -80°C was also tested. Again, the protein was denatured when thawed from either temperature. Similarly, RpiRa CTD₇₉₋₂₆₆ is not stable to high temperature and denatures when exposed to temperatures above 30°C for overnight (Figure 9.9). RpiRa CTD₇₉₋₂₆₆ is also not stable to lyophilization. Lyophilization of the protein from (NH₄)₂CO₃ or a phosphate buffer results in a precipitated protein that will not re-dissolve into the original buffer. In practice, RpiRa CTD₇₉₋₂₆₆ was always kept in room temperature in the optimized NMR buffer.

There are two common methods to perform buffer exchange: dialysis and ultrafiltration using a centrifuge. Both float-A-analyzer and traditional dialysis membrane tubing were tested and show comparable results. Importantly, a dialysis approach to a buffer exchange outperformed Amicon ultrafiltration. Amicon ultrafiltration caused a significant amount of protein purification probably because it is a harsh method relative to dialysis.

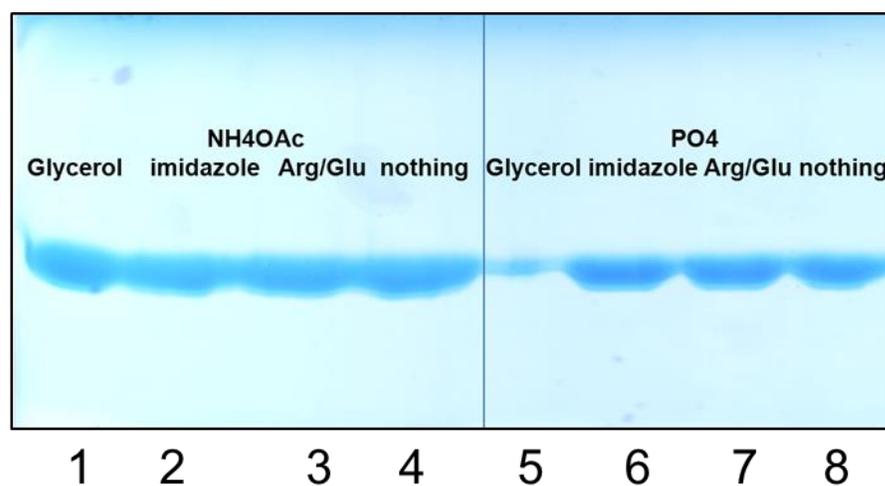


Figure 9.7 The impact of adding stability agents to the RpiRa CTD₇₉₋₂₆₆ NMR buffer. Two optimal NMR buffer systems were tested for adding these additives. No observable changes in the RpiRa CTD₇₉₋₂₆₆ samples were observed after weeks of storage.

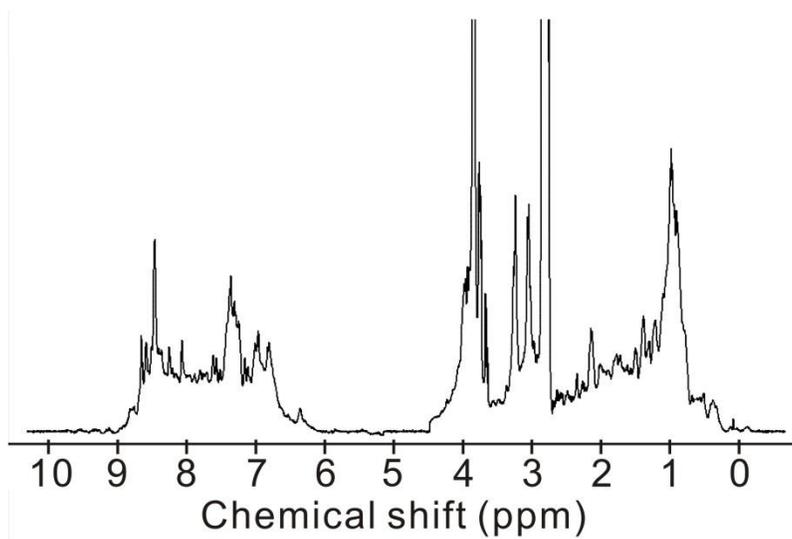


Figure 9.8 The 1D ^1H NMR spectrum of RpiRa CTD₇₉₋₂₆₆. The water peak in the region of 4.6-5.2 ppm was removed.

9.3.6 Isotopic labeling of RpiRa CTD

Uniformly ^{15}N and ^{13}C labeled RpiRa CTD₇₉₋₂₆₆ was produced using minimal medium supplemented with ^{15}N NH_4Cl and ^{13}C glucose. Deuteration of the aromatic and aliphatic protons in the protein can be achieved by expressing the protein in D_2O .⁴⁷ Deuteration of the non-exchangeable proton was shown to improve the resolution and sensitivity of NMR experiments by reducing the overall relaxation rates of NMR active nuclei (Figure 9.10a and 9.10b).^{48, 29} The difficulty of deuterium incorporation was demonstrated by the fact that increasing the percentage of D_2O in the cell culture media dramatically slowed cell growth. 80% D_2O was shown to be the highest percentage of

D₂O that the *E. coli* strain could tolerate. Although, after adaptation to D₂O, the *E. coli* growth was comparable to cell growth in H₂O.

The non-exchangeable and exchangeable protons are both labeled with deuterium during the cell culture, so it takes additional effort to exchange deuterated amides back to protons to be observable by NMR. It may be particularly challenging to exchange amides buried in the core of the protein or involved in secondary structures. Therefore, it may require an increase in the temperature to partially unfold the protein in order to expose all the amides to the solvent and increase the HD exchange rate. Also, it may simply require waiting days or weeks until all of the deuterated amides have effectively changed back to protons. As shown in Figure 9.8, the peaks of the 2D ¹H, ¹⁵N HSQC spectrum were sharpened after an overnight incubation at 30°C in H₂O. But a longer incubation time leads to the denaturing of RpiRa CTD₇₉₋₂₆₆. An alternative and safer approach was chosen in which the protein was placed in a 50 mL buffer at pH 7.4 for 2-4 days at room temperature. This was effective in back exchange all of the amides since the amide exchange rate is higher at higher pH. After the amide exchange was completed, the 2D ¹H, ¹⁵N HSQC spectrum improved and more peaks were observable (Figure 9.10a and b). An alternative approach is to increase the sample temperature during NMR data collection. The 2D ¹H, ¹⁵N HSQC spectrum was collected at 25°C, 27°C, 29°C, and 31°C. Temperatures higher than 31°C caused significant protein precipitation. No improvement in the 2D ¹H, ¹⁵N HSQC spectrum was observed, so the sample temperature was kept at 25°C for all NMR experiments. Figures 9.10b-c illustrates an interesting phenomenon. Over a month time frame, the RpiRa CTD₇₉₋₂₆₆ NMR sample slowly precipitated and then

eventually stabilized. This resulted in a higher quality 2D ^1H , ^{15}N HSQC spectrum, suggesting that there might be a concentration dependent monomer-dimer transition.

RpiRa CTD₇₉₋₂₆₆ is a relatively large protein for NMR, so using TROSY experiment are expected to improve the spectral quality by generating sharper peaks. As shown in Figure 9.10d, the collection of a 2D ^1H , ^{15}N TROSY-HSQC experiment on a 700 MHz NMR spectrometer resulted in a dramatic improvement in the quality of the NMR spectrum compared to the other 2D ^1H , ^{15}N HSQC spectra.

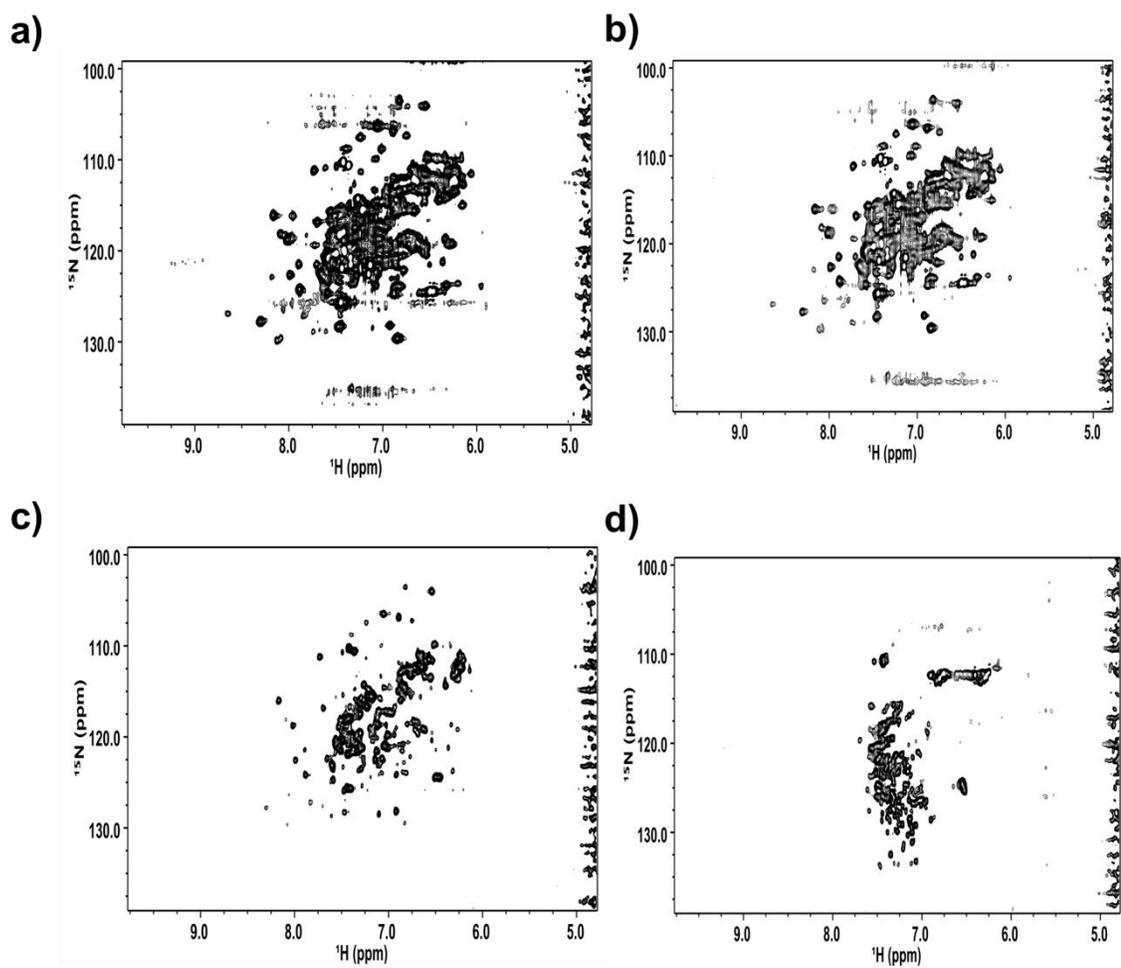


Figure 9.9 The overnight incubation of an RpiRa NMR sample at a) 25°C; b) 30°C; c) 35°C; d) 40°C.

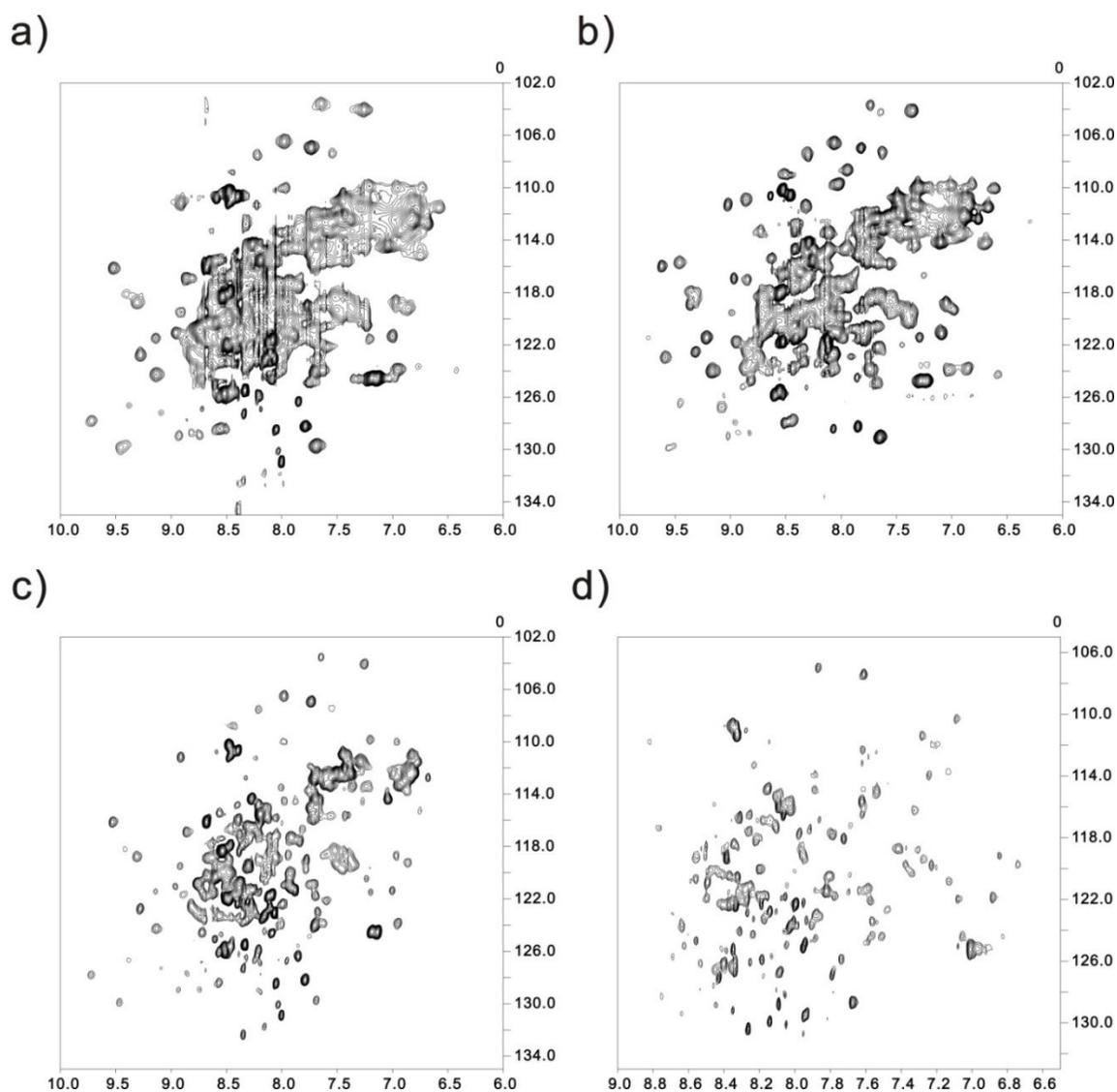


Figure 9.10 a) 2D ^1H , ^{15}N HSQC spectrum of uniformly ^{15}N labeled RpiRa CTD₇₉₋₂₆₆ ; b) 2D ^1H , ^{15}N HSQC spectrum of uniformly ^2H (80%), ^{15}N labeled RpiRa CTD₇₉₋₂₆₆; c) 2D ^1H , ^{15}N HSQC spectrum of uniformly ^2H (80%), ^{15}N labeled RpiRa CTD₇₉₋₂₆₆ sample from (b), but after a month; d) 2D ^1H , ^{15}N TROSY-HSQC spectrum of uniformly ^2H (80%), ^{15}N labeled RpiRa CTD₇₉₋₂₆₆.

9.4 Discussion

The optimization of experimental conditions for the overexpression, purification and storage of RpiRa CTD₇₉₋₂₆₆ was primarily based on “trial and error” due to the unpredictable nature of proteins. This initial and essential work was done to achieve an RpiRa CTD₇₉₋₂₆₆ sample amenable for future NMR structural biology studies. This effort has led to the identification of an appropriate expression system and purification protocols, to an optimal buffer condition to maintain protein stability and obtain high quality NMR spectra, and an effective isotope labeling strategy to enable the use of TROSY-based experiments. Further work in the application of selective labeling and non-uniform labeling are still required to improve the likelihood that the NMR backbone assignments can be obtained for RpiRa CTD₇₉₋₂₆₆.

9.5 References

- (1) Zhu, Y.; Weiss, E. C.; Otto, M.; Fey, P. D.; Smeltzer, M. S.; Somerville, G. A. *Infect Immun* **2007**, *75*, 4219.
- (2) Sadykov, M. R.; Olson, M. E.; Halouska, S.; Zhu, Y.; Fey, P. D.; Powers, R.; Somerville, G. A. *J Bacteriol* **2008**, *190*, 7621.
- (3) Zhang, B.; Powers, R. *Future Med Chem* **2012**, *4*, 1273.
- (4) Zhu, Y.; Nandakumar, R.; Sadykov, M. R.; Madayiputhiya, N.; Luong, T. T.; Gaupp, R.; Lee, C. Y.; Somerville, G. A. *J Bacteriol* **2011**, *193*, 6187.
- (5) Somerville, G. A.; Proctor, R. A. *Microbiol Mol Biol Rev* **2009**, *73*, 233.

- (6) Mailloux, R. J.; B ériault, R.; Lemire, J.; Singh, R.; Ch énier, D. R.; Hamel, R. D.; Appanna, V. D. *PLoS One* **2007**, *2*, e690.
- (7) Sadykov, M. R.; Mattes, T. A.; Luong, T. T.; Zhu, Y.; Day, S. R.; Sifri, C. D.; Lee, C. Y.; Somerville, G. A. *J Bacteriol* **2010**, *192*, 1459.
- (8) Lopilato, J. E.; Garwin, J. L.; Emr, S. D.; Silhavy, T. J.; Beckwith, J. R. *J Bacteriol* **1984**, *158*, 665.
- (9) Sorensen, K. I.; Hove-Jensen, B. *J Bacteriol* **1996**, *178*, 1003.
- (10) Edwards, T.; Abramov, A.; Smith, E.; Baydo, R.; Leonard, J.; Leibly, D.; Thompkins, K.; Clifton, M.; Gardberg, A.; Staker, B.; Van Voorhis, W.; Myler, P.; Stewart, L. *BMC Structural Biology* **2011**, *11*, 39.
- (11) Zhang, R.; Andersson, C. E.; Savchenko, A.; Skarina, T.; Evdokimova, E.; Beasley, S.; Arrowsmith, C. H.; Edwards, A. M.; Joachimiak, A.; Mowbray, S. L. *Structure* **2003**, *11*, 31.
- (12) Essenberg, M. K.; Cooper, R. A. *Eur J Biochem* **1975**, *55*, 323.
- (13) Roos, A. K.; Mariano, S.; Kowalinski, E.; Salmon, L.; Mowbray, S. L. *J Mol Biol* **2008**, *382*, 667.
- (14) Daddaoua, A.; Krell, T.; Ramos, J. L. *J Biol Chem* **2009**, *284*, 21360.
- (15) Yamamoto, H.; Serizawa, M.; Thompson, J.; Sekiguchi, J. *J Bacteriol* **2001**, *183*, 5110.
- (16) Bateman, A. *Trends in Biochemical Sciences* **1999**, *24*, 94.
- (17) Baneyx, F. *Curr Opin Biotechnol* **1999**, *10*, 411.

- (18) Tegel, H.; Steen, J.; Konrad, A.; Nikdin, H.; Pettersson, K.; Stenvall, M.; Tourle, S.; Wrethagen, U.; Xu, L.; Yderland, L.; Uhlen, M.; Hober, S.; Ottosson, J. *Biotechnol J* **2009**, *4*, 51.
- (19) Prinz, W. A.; Aslund, F.; Holmgren, A.; Beckwith, J. *J Biol Chem* **1997**, *272*, 15661.
- (20) Guzman, L. M.; Belin, D.; Carson, M. J.; Beckwith, J. *J Bacteriol* **1995**, *177*, 4121.
- (21) Bessette, P. H.; Aslund, F.; Beckwith, J.; Georgiou, G. *Proc Natl Acad Sci U S A* **1999**, *96*, 13703.
- (22) Brown, B. L.; Grigoriu, S.; Kim, Y.; Arruda, J. M.; Davenport, A.; Wood, T. K.; Peti, W.; Page, R. *PLoS Pathog* **2009**, *5*, e1000706.
- (23) Terpe, K. *Appl Microbiol Biotechnol* **2003**, *60*, 523.
- (24) Lichty, J. J.; Malecki, J. L.; Agnew, H. D.; Michelson-Horowitz, D. J.; Tan, S. *Protein Expr Purif* **2005**, *41*, 98.
- (25) Walls, D.; Loughran, S. T. *Methods Mol Biol* **2011**, *681*, 151.
- (26) Rossi, P.; Swapna, G. V.; Huang, Y. J.; Aramini, J. M.; Anklin, C.; Conover, K.; Hamilton, K.; Xiao, R.; Acton, T. B.; Ertekin, A.; Everett, J. K.; Montelione, G. T. *J Biomol NMR* **2010**, *46*, 11.
- (27) Cavanagh, J.; Fairbrother, W.; Palmer, A. I.; Skelton, N.; Rance, M. *Protein NMR Spectroscopy: Principles and Practice*; Academic Press: Waltham, Massachusetts, 2006.
- (28) Mercier, K. A.; Baran, M.; Ramanathan, V.; Revesz, P.; Xiao, R.; Montelione, G. T.; Powers, R. *J Am Chem Soc* **2006**, *128*, 15292.

- (29) Yamazaki, T.; Lee, W.; Arrowsmith, C. H.; Muhandiram, D. R.; Kay, L. E. *Journal of the American Chemical Society* **1994**, *116*, 11655.
- (30) Lee, K. M.; Androphy, E. J.; Baleja, J. D. *J Biomol NMR* **1995**, *5*, 93.
- (31) Fernandez, C.; Wider, G. *Curr Opin Struct Biol* **2003**, *13*, 570.
- (32) Eletsky, A.; Kienhöfer, A.; Pervushin, K. *Journal of Biomolecular NMR* **2001**, *20*, 177.
- (33) LeMaster, D. M. *Annual Review of Biophysics and Biophysical Chemistry* **1990**, *19*, 243.
- (34) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J Biomol NMR* **1995**, *6*, 277.
- (35) McGuffin, L. J.; Bryson, K.; Jones, D. T. *Bioinformatics* **2000**, *16*, 404.
- (36) Quevillon, E.; Silventoinen, V.; Pillai, S.; Harte, N.; Mulder, N.; Apweiler, R.; Lopez, R. *Nucleic Acids Res* **2005**, *33*, W116.
- (37) Cheng, J. *Nucleic Acids Res* **2007**, *35*, W354.
- (38) Buchan, D. W. A.; Minneci, F.; Nugent, T. C. O.; Bryson, K.; Jones, D. T. *Nucleic Acids Research* **2013**, *41*, W349.
- (39) Arnold, K.; Bordoli, L.; Kopp, J.; Schwede, T. *Bioinformatics* **2006**, *22*, 195.
- (40) Rost, B.; Yachdav, G.; Liu, J. *Nucleic Acids Res* **2004**, *32*, W321.
- (41) Jones, P.; Binns, D.; Chang, H.-Y.; Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; Maslen, J.; Mitchell, A.; Nuka, G.; Pesseat, S.; Quinn, A. F.; Sangrador-Vegas, A.; Scheremetjew, M.; Yong, S.-Y.; Lopez, R.; Hunter, S. *Bioinformatics* **2014**.

- (42) Gasteiger E., H. C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A. *Protein Identification and Analysis Tools on the ExPASy Server*; Humana Press: New York, 2005.
- (43) Sivashanmugam, A.; Murray, V.; Cui, C.; Zhang, Y.; Wang, J.; Li, Q. *Protein Sci* **2009**, *18*, 936.
- (44) Soini, J.; Ukkonen, K.; Neubauer, P. *Microbial Cell Factories* **2008**, *7*, 26.
- (45) Tan, S. C.; Yiap, B. C. *Journal of Biomedicine and Biotechnology* **2009**, *2009*.
- (46) Pielak, G. J.; Li, C.; Miklos, A. C.; Schlesinger, A. P.; Slade, K. M.; Wang, G. F.; Zigoneanu, I. G. *Biochemistry* **2009**, *48*, 226.
- (47) Markus, M. A.; Dayie, K. T.; Matsudaira, P.; Wagner, G. *Journal of Magnetic Resonance, Series B* **1994**, *105*, 192.
- (48) Leiting, B.; Marsilio, F.; O'Connell, J. F. *Anal Biochem* **1998**, *265*, 351.

APPENDIX B: Medium and buffer recipes

Note: All the buffers and media were prepared in nanopure H₂O and autoclaved.

1. LB media (1 L):

10.0 g Tryptone

5.0 g Yeast extract

10.0 g NaCl

2. LB media agar plate (1 L):

10.0 g Tryptone

5.0 g Yeast extract

10.0 g NaCl

15.0 g Agar

3. M9 minimal medium (1 L):

A) M9 salt (1 L, 5X)

64 g $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$

15.0 g KH_2PO_4

2.5g NaCl

5.0 g NH_4Cl (or $^{15}\text{NH}_4\text{Cl}$)

0.2 mL of 10 mM FeCl_3 stock solution

B) M9 minimal medium (1L)

200 mL 5XM9 salt

2.0 mL 1M MgSO_4 stock solution

2.0 mL 1M CaCl_2 stock solution

2 g glucose (or ^{13}C -glucose)

4. High-cell-density medium (1 L):

13.4 g $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$

3.4 g KH_2PO_4

0.58 g NaCl

5 mL 1 M MgSO₄ stock solution

0.2 CaCl₂ stock solution

1.0 g NH₄Cl (or ¹⁵NH₄Cl)

10 g glucose (or ¹³C-glucose)

5. Extraction buffer (1 L):

2.63 g KH₂PO₄

5.34 g K₂HPO₄

17.55 g NaCl

50 mL glycerol

6. Wash buffer (1 L):

2.63 g KH₂PO₄

5.34 g K₂HPO₄

680.8 g NaCl

680.8 mg imidazole

7. Elution buffer (1 L):

2.63 g KH₂PO₄

5.34 g K₂HPO₄

3.40 g NaCl

3.40 g imidazole (for 50 mM) or 10.2 g imidazole (for 150 mM)

8. NMR sample buffer (1 L):

8.1 Potassium phosphate buffer saline^{*}

2.72 g KH_2PO_4

5.85 g NaCl

1.54 g DTT (or 0.57 g TCEP-HCl)

0.20 g NaN_3

^{*}adjust pH to 4.5

8.2 Ammonium acetate buffer^o

1.54 g NH_4OAc

5.85 g NaCl

1.54 g DTT (or 0.57 g TCEP-HCl)

0.20 g NaN_3

^oadjust pH to 4.5

CHAPTER 10

SUMMARY AND FUTURE DIRECTIONS

10.1 Summary

Simplistically, metabolomics can be considered as a continuous array of metabolite assays, but it is not feasible to optimize a single experimental condition to accurately measure each metabolite in this hypothetical array of assays. Therefore subtle differences in experimental procedures may lead to detectable changes in metabolite concentrations. This is an inherent challenge of metabolomics, but this issue is also significantly diminished by relying on a comparative (instead of an absolute) analysis between two or more samples. As long as experimental consistency is maintained between all samples, the underlying errors will be constant. Thus, these errors will be irrelevant and will “subtract out” when two or more samples are compared to identify “differences”. So the first aim of my thesis was to exhaustively evaluate all of the experimental factors that may contribute to any sample inconsistency and negatively impact the outcome of an NMR-based metabolomics study.

Maintaining consistency across a dataset of 1D ^1H NMR spectra is critical for an untargeted metabolomics study because multivariate statistical techniques such as PCA or OPLS-DA are intrinsically sensitive to all variations in the data, regardless of the origin of these differences. Thus, biologically irrelevant variations between classes that are introduced by the investigator or by the experimental protocols need to be removed in order to avoid misinterpretations and erroneous conclusions. Towards this end, we

optimized sample preparation protocols for bacteria and mammalian cell lines to maximize reproducibility between replicates, and to achieve complete and efficient extraction of the cellular metabolome. These protocols are described in detail in Chapters 3 and 4. The entire process includes: 1) growing and harvesting bacterial cells, 2) extracting the metabolome, 3) NMR data collection, processing and analysis, 4) statistical analysis, 5) metabolite identification, and 6) metabolic network generation. Consider, for example, the preparation of a metabolomics sample, we were able to identify and correct for a number of variables that decrease consistency between replicates. We simultaneously use three different normalization schemes to correct for variations in total metabolite concentration: the number of cells per sample, protein concentration per sample, and Standard Normal Variate (SNV) data preprocessing. We can harvest and quench each metabolomics sample in less than 20 seconds using centrifugation or filtration. The type of glass beads, the bead beating program, and the temperature used for cell lysis were all optimized to obtain complete extraction of the cellular metabolome. Similarly, the type of solvent and buffer to obtain complete extraction of the metabolome was also optimized. NMR experimental parameters that include the number of scan, number of data points, receiver gain settings, water suppression pulse program, spectral width, and relaxation time were optimized for each project. These validated experimental protocols were successfully applied to all metabolic projects. Based on this exhaustive analysis, we have a thorough understanding of the experimental parameters and conditions that significantly impact the quality and outcome of a metabolomics study. Consequently, in future studies, we only need to focus on these specific steps to

successfully optimize the preparation and analysis of metabolomics samples. In return, this will help increase the efficiency and consistency of future metabolomics studies.

The second aim of my thesis was to understand the regulatory role that the TCA cycle plays in *S. aureus* and *S. epidermidis* biofilm formation. Specifically, how do bacteria respond to a disparate set of environmental stress conditions that are all known to induce biofilm formation? Iron-depletion, the addition of ethanol, glucose or low-dosages of antibiotics, temperature, and anaerobic conditions are all known to induce biofilm formation. We observed that iron-depletion, the addition of ethanol, glucose, low-dose tetracycline, and anaerobic conditions perturbed the metabolome of *S. epidermidis* in a manner similar to an aconitase mutant. A metabolomics 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, which is a major precursor to biofilm formation. ¹³C glucose was used to trace carbon flow in *S. aureus* and *S. epidermidis* under conditions that induce biofilm formation. The observed inactivation of the TCA cycle was correlated with an increase in a number of precursors of PIA, which indicates PIA biosynthesis is reversely regulated by TCA cycle.¹ As a result of these observations, the TCA cycle can be considered as a metabolic sensor of environmental stressors. In summary, the metabolomics study of the role that TCA cycle plays in mediating intracellular energy and homeostasis in response to environmental stressors significantly enhanced our understanding of a metabolism-dependent mechanism of biofilm formation.

NMR-based metabolomics was also employed to investigate the impact of TCA cycle activity on the *S. aureus* metabolome with a specific emphasis on the role of CcpE, a putative LysR-type of regulator and a predicted activator of the TCA cycle. The construction of a metabolic network based on perturbations in CcpE activity indicated that changes in the *S. aureus* metabolome was centered on the TCA cycle. This clearly supported our hypothesis that CcpE functions as a positive regulator of the TCA cycle. Additionally, our analysis suggested that *citB* is a target of CcpE. This study combined with previous studies of AcnA, CcpA, RpiR and CodY as metabolic regulators, describe a group of genes in staphylococci that control the function of TCA cycle. Another metabolomic study in *S. aureus* was focused on the influence of oxygen and iron limitation. By decreasing the flask-to-medium ratio, the *S. aureus* metabolome shifts from oxidative to a heterofermentative metabolism. Similarly, iron-limitation also shifted the metabolic profile to a heterofermentative metabolism. Both stressors caused a suppression of the TCA cycle, as further evident by the decreased transcription of aconitase (*citB/acnA*). Importantly, the effects of iron-limitation were obscured by the oxygen-limitation. Once again, different environmental stressors were shown to have a similar effect on the central metabolism in staphylococci, but our results also demonstrate that two specific environmental stressors are connected and interdependent.

The last aim of my thesis was the optimization of experimental conditions for the overexpression, purification and storage of RpiRa CTD₇₉₋₂₆₆. This initial and essential work was done to achieve an RpiRa CTD₇₉₋₂₆₆ sample amenable for future NMR structural biology studies. This effort has led to the identification of an appropriate T7

expression system and the selection of the correct sequence for the RpiRa CTD₇₉₋₂₆₆. Four different growth strategies were compared and the aerobic growth on high cell density medium was chosen to maximize the protein yield. A protein purification protocol and NMR sample condition was also optimized. The optimal buffer condition consisted of a phosphate buffer pH 4.5 with 2 mM TCEP-HCl, 100 mM NaCl and 0.02 % NaN₃ was chosen to maintain protein stability and obtain high quality NMR spectra. An effective ²H, ¹³C, ¹⁵N triple-labeling strategy to enable the use of TROSY-based experiments was also implemented.

10.2 Future directions

The protocol used to extract and prepare a metabolomics sample also influences the choice of analytical method (MS vs. NMR) and the corresponding analysis strategy. There is a growing consensus that using multiple analytical techniques is a preferred approach to thoroughly characterize the metabolome. This, in turn, requires the integration of different methods in terms of sample handling, instrumental analysis, and data process. In regards to NMR-based metabolomics, additional experimental techniques for selective detection and suppression should also be considered. For example, ³¹P NMR and ¹⁵N NMR can be used to highlight groups of metabolite such as phosphorylated sugars and amino acids.² ³¹P and ¹⁵N isotopic labeling coupled with monitoring the flow of nitrogen and phosphorus throughout the metabolome may yield interesting insights about metabolic pathways not observed using ¹³C-labeled metabolites. Analyzing a

subset of metabolites based on prior knowledge of biological relevance is a unique feature of NMR. Investigating a smaller set of metabolites would also simplify metabolite identification since the corresponding NMR spectra are greatly simplified and peak overlap is significantly diminished. This addresses a major bottleneck of NMR metabolomics – the manual peak assignments. Currently, metabolite identification is based on the manual analysis of NMR spectra that requires using the entirety of the spectral information combined with cross-validation using multiple database, all available biological information, the experience of the investigator, and an iterative analysis that relies on informed and educated guesses. Ideally, this process should be automated by using a combination of new software and a series of NMR experiments such as 2D COSY, TOCSY, HSQC, HMBC, and HSQC-TOCSY.

Staphylococcus epidermidis and *Staphylococcus aureus* are the leading causes of hospital-associated infections.³ The only remaining treatment for most staphylococci is glycopeptide antibiotic vancomycin, although a vancomycin resistant strain has been found.⁴ Poor antibiotic penetration, nutrient limitation, slow growth, adaptive stress responses, and formation of persister cells are all hypothesized to constitute a defense system.⁵ Many of these mechanisms are metabolism associated. A critical fact is that although the type and severity of the virulence produced, or level of drug resistance by these bacteria can vary, the metabolic pathways that they utilize are very similar. Staphylococcal species can synthesize all their virulence factors based on 13 small molecules derived from central metabolic pathways.⁶ A study showed that over 90% of the enzymes, metabolic reactions and metabolites were common to all strains.⁷ A

common mechanism of their drug-resistance is biofilm-related infectious manifestations associated with implanted biomaterials and biofilm formation can be largely influenced by central metabolism.⁸ It is appealing to study the common feature of the metabolism so as to not only understand their pathogenesis development, but also to search for next generation antibiotics.^{6,1}

In the structural characterization of RpiRa CTD, the most important and long term goal is to solve its three dimensional NMR structure. Other aspects of interests include: (i) different labeling strategies that highlight selected residues to simplify the NMR assignments, (ii) implementation of non-uniform sampling to increase spectrum quality to enable determining the NMR structure, (iii) ligand screening (FAST-NMR) that identifies small molecules potentially related to the function of RpiRa and the identification of the functionally relevant ligand binding site, and (vi) additional molecular and cellular biology experiments to understand the regulation mechanism of RpiRa and its role in the pentose phosphate pathway.

10.3 References

- (1) Sadykov, M. R.; Olson, M. E.; Halouska, S.; Zhu, Y.; Fey, P. D.; Powers, R.; Somerville, G. A. *J Bacteriol* **2008**, *190*, 7621.
- (2) Gowda, G. A.; Shanaiah, N.; Raftery, D. *Adv Exp Med Biol* **2012**, *992*, 147.
- (3) von Eiff, C.; Peters, G.; Heilmann, C. *Lancet Infect Dis* **2002**, *2*, 677.

- (4) Gill, S. R.; Fouts, D. E.; Archer, G. L.; Mongodin, E. F.; Deboy, R. T.; Ravel, J.; Paulsen, I. T.; Kolonay, J. F.; Brinkac, L.; Beanan, M.; Dodson, R. J.; Daugherty, S. C.; Madupu, R.; Angiuoli, S. V.; Durkin, A. S.; Haft, D. H.; Vamathevan, J.; Khouri, H.; Utterback, T.; Lee, C.; Dimitrov, G.; Jiang, L.; Qin, H.; Weidman, J.; Tran, K.; Kang, K.; Hance, I. R.; Nelson, K. E.; Fraser, C. M. *J Bacteriol* **2005**, *187*, 2426.
- (5) Stewart, P. S. *Int J Med Microbiol* **2002**, *292*, 107.
- (6) Somerville, G. A.; Proctor, R. A. *Microbiol Mol Biol Rev* **2009**, *73*, 233.
- (7) Lee, D. S.; Burd, H.; Liu, J.; Almaas, E.; Wiest, O.; Barabasi, A. L.; Oltvai, Z. N.; Kapatral, V. *J Bacteriol* **2009**, *191*, 4015.
- (8) Krimmer, V.; Merkert, H.; von Eiff, C.; Frosch, M.; Eulert, J.; Lohr, J. F.; Hacker, J.; Ziebuhr, W. *J Clin Microbiol* **1999**, *37*, 2667.