University of Alberta

Comparative proteomic and metabolomic analysis of a Staphylococcus variant (SG1) cultured in the presence and absence of butanol

by

Feifei Fu

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science

Department of Chemistry

Fall 2013
Edmonton, Alberta

Abstract

My research focuses on proteomic and metabolomic analysis of a solvent tolerant strain of *Staphylococcus warneri* called SG1 cultured in the presence and absence of 1-Butanol.

On the proteomic analysis, the tryptic digests of SG1 were either directly analyzed by two-dimensional liquid chromatography electrospray ionization quadrupole time-of-flight mass spectrometry (2D-LC-ESI-QTOF MS) for protein identification or isotope labeled using dimethylation after guanidination (2-MEGA) followed by 2D-LC-ESI-QTOF MS for relative protein quantification.

On the metabolomic analysis, the extracted metabolites were either labeled with $^{12}C/^{13}C$ dansyl chloride for relative quantitation of amine- and phenol-containing metabolites or labeled with $^{12}C/^{13}C$ *p*-dimethylaminophenacyl (DmPA) bromide for relative quantitation of carboxylic acid-containing metabolites.

Finally, proteomic data and metabolomic data were combined and compared to help elucidate the solvent tolerant mechanism of SG1. The thesis work highlights the potential and significance of combining proteomic and metabolomic analyses for studying biological systems.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

2D	two dimensional
2MEGA	dimethylation after guanidination
AQUA	absolute quantification
$BtOH^-$	without 1-Butanol
$BtOH^+$	with 1.5 % 1-Butanol
CID	collision induced dissociation
COG	cluster of orthologous groups
DC	direct current
DIL	differential isotope labeling
DmPA	*p*-dimethylaminophenacyl
ESI	electrospray ionization
FT-ICR	Fourier transform ion cyclotron resonance
GC	gas chromatography
GRAVY	grand average of hydropathy
HILIC	hydrophilic interaction chromatography
IEF	isoelectric focussing
ICAT	isotope coded affinity tag
iTRAQ	isobaric tags for relative and absolute quantification
KEGG	kyoto encyclopedia of genes and genome
m/z	mass to charge ratio
MALDI	matrix-assisted laser desorption ionization
MRM	multiple reaction monitoring
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MudPIT	multidimensional protein identification technology
NMR	nuclear magnetic resonance
PSMs	peptide-spectrum matches
QTOF	quadrupole time-of-flight
RF	radio frequency
RP	reversed phase

S/N	signal-to-noise
SCX	strong cation exchange
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SILAC	stable isotope labelling by amino acids in cell culture
TCA	trichloroacetic acid
TCA cycle	tricarboxylic acid cycle
TFA	trifluoroacetic acid

Chapter 1. Introduction

Proteomics is a newly emerging field that has proven to be very important for many research areas of biology and medical science[1]. The term "proteome" was first coined by Dr. Marc Wilkins to refer a set of proteins encoded by the genome[2]. And proteomics, the study of proteome, now evokes all the proteins in any living organism, including all protein isoforms, their post-translational modifications, protein-protein interactions, protein structures and their complexes in higher order[3]. The most common and powerful tool in proteomics is mass spectrometry (MS) which utilizes mass analysis for identification of proteins and can provide massive information about the proteome of interest.

Metabolomics is the study of endogenous small molecules (metabolites) in biosystems. The term "metabolome" was first coined in 1998 by Oliver[4]. With effective chromatography separation, mass spectrometry offers the ability to profile and quantify individual compounds in biological samples with high sensitivity and specificity. Thus, liquid chromatography mass spectrometry (LC-MS) has become one of the most important tools in metabolomic analysis[5,6]. Due to the vast diversity of metabolites, targeted methods analyzing a group of metabolites sharing similar structural moieties are often used, increasing the possibility of detecting more metabolites[7,8] .

My thesis focuses on studying the proteome and the metabolome of an interesting butanol-tolerant Gram-positive *Staphylococcus warneri* strain SG1. My goal is to assemble the proteome profile for this bacterium and quantify the differentially expressed proteins in the presence and absence of 1-Butanol. In addition, we would like to observe and quantify the changes in metabolite levels when SG1 is subjected to butanol challenge. To achieve this goal, we applied several types of MS technologies for the proteomic and metabolomic studies. Due to the scope of this chapter and the fact that there are numerous excellent reviews

on MS for proteomics and metabolomics, I will not cover all the areas in this chapter, but rather, focus on the discussion of the most relevant aspects of the topics to my thesis work. First, I will introduce the standard sample preparation approach for bottom-up proteomics (shotgun proteomics). Second, the MS instrumentation including common ionization techniques, quadrupole time-of-flight (QTOF) MS and Fourier transform ion cyclone resonance (FTICR) MS will be introduced, followed by an overview of various quantitative proteomic approaches. Finally, I will introduce the quantitative metabolomic approaches and provide the scope of my thesis work.

1.1 Overview of MS-based bottom-up proteomic analysis

A proteomic study begins with the sample preparation in which proteins are either enzymatically digested into peptides (bottom-up approach)[9,10] or analyzed in intact form (top-down approach)[11].

When tackling the highly complex samples for a large scale investigation of a proteome, the bottom-up approach is the most popular method. Bottom-up proteomics is also called shotgun proteomics[12,13], an approach where the proteins are proteolytically digested into peptides before MS analysis and the acquired peptide masses and sequences are used for identification of the corresponding proteins. Most applications of shotgun proteomics require tandem MS acquisition where the peptides are further fragmented by collision induced dissociation (CID). The most widely used method for protein identification is based on database search where experimental data (i.e., MS/MS spectrum) is compared with the predicted *in silico* fragmentation patterns of the peptide of interest[14,15]. The advantages of the bottom-up approach include the better front-end separation of peptides, higher sensitivity and higher throughput than top-down method. On the other hand, the drawbacks of the bottom-up approach include limited protein sequence coverage by identified peptides, loss of labile post-translational modifications, and ambiguity of the origin for redundant peptide sequences[16]. In

the following section, I will discuss the critical steps for sample preparation including peptide digestion and liquid chromatography (LC) separation of peptides for MS analysis.

1.1.1 Protein extraction and digestion

In proteomics experiments, the first step is to extract proteins from the biological samples such as cells, tissues or biofluids. Protein extraction is the first critical step for the success of the whole experiment. It can be achieved by physical approaches that rupture the cellular structure by physical actions (pressure, sonication, freeze/thaw cycles, etc.), or by extraction buffers containing enzymes (e.g. lysozyme) or chemicals (e.g. surfactants) that disrupt the cell membrane or by the combination of both. In addition to proteins, extracts from complex biological samples often contains other types of cellular components, such as lipids, salts, nucleic acids and other macromolecules, which may interfere with MS analysis. To remove those undesired components, protein precipitation by solvent such as acetone[17] and trichloroacetic acid (TCA) is routinely used.

Once proteins are enriched from the extract by protein precipitation, they are solubilized in a solution and further subjected to chemical fragmentation or enzymatic digestion, forming smaller peptides prior to chromatographic separation. There are only a few chemicals available for chemical fragmentation of proteins with acceptable specificity. One example is CNBr which cleaves amide bond at the C-terminus side of methionine. Other enzymes such as trypsin, chymotrypsin, Lys-C, Lys-N and Glu-C/V8 are more commonly used to perform enzymatic digestion due to the availability of enzymes and the better specificity of digestion. Among them, trypsin is one of the most widely used enzymes for protein digestion. It cleaves peptide bonds to give peptides with arginine or lysine at the C-terminus, except when either is followed by proline. The specificity of trypsin is very high, and moreover, the trypsin-generated peptides (tryptic peptides) typically have mass range from 600 to 3000 Da which is ideal for MS

acquisition. Another advantage of utilizing trypsin come to the fact that peptides containing arginine and lysine at the C-termini are more efficient to be protonated and ionized in electrospray ionization (ESI) or matrix-assisted laser desorption ionization (MALDI), compared to those without arginine and lysine at the C-terminus.

1.1.2 Liquid chromatography (LC) Separation

When analyzing complex samples, hundreds of thousands of peptides are generated from protein digestion. Analyzing such a complex mixture of peptides by MS has been proved to be a challenge. There are mainly two reasons. First, the mass window of current state-of-art MS used to isolate the peptide ions for fragmentation needs to be sufficiently wide (2-4 Da) to provide high sensitivity detection of the fragment ions, and this mass window can contain multiple peptides which would make the interpretation of the resultant MS/MS spectra very difficult. Second, the ionization efficiencies of peptides are very different. Peptides that are in low concentrations or have low ionization efficiencies are very likely suppressed while only a few peptides could be ionized and detected by MS. Thus, separation of peptides to reduce complexity of samples is required prior to MS detection.

For peptide separation, the most widely used technique is high performance liquid chromatography (HPLC or LC)[18]. LC separation is achieved based on the difference in interaction of peptides with a stationary phase packed in a column. The materials of stationary phases determine the surface chemistry of the stationary phase and the separation mechanism. There are two LC separation techniques commonly used in proteomics. The first one is reversed phase LC (RPLC) where the stationary phase contains a non-polar functional group such as C18 and the retention of peptides is facilitated by hydrophobic interaction with the stationary phase. Peptide elution is done using two-component solvent system with increasing proportion of non-polar component (e.g., increasing the content of

acetonitrile). The second technique is strong cation exchange (SCX). In SCX, the stationary phase is made of materials containing anionic function group (e.g. -SO_3^-) on the surface. Prior to loading into SCX LC column, the pH of the peptides solution is adjusted to acidic so that most peptides are protonated and positively charged which leads to retention in SCX LC column by ionic interaction with the stationary phase after loading. Peptides are eluted by increasing the concentration of cations in the solvent mixture (such as increasing KCl concentration or increasing the pH).

In proteomic experiments, LC separation is usually coupled with MS detection. And in most case, RPLC is used since RPLC uses compatible buffers with ESI and MALDI and it can provide higher separation efficiency compared to SCX. However, MS coupled to RPLC alone is not able to resolve most of the complex proteomic samples, and thus multi-dimensional LC is required[18,19]. In principle, the more dimensions of LC separation used the better separation efficiency and MS detection. But in reality, increasing the dimension of LC separation significantly increases the demand of labor and time. Therefore, two dimensions of LC separation are used in most cases of shotgun proteomics where the SCX separation serves as the first dimension and RPLC serves as the second one. We call such a technology as 2D-LC MS. These two LC techniques are considered orthogonal since SCX separation is based on the difference in ionic interaction while RPLC separation is based on the difference in hydrophobicity of peptides. Thus peptides with similar retention time in one LC column are very likely resolved in the other LC column.

There are two instrumental configurations for 2D-LC MS, on-line 2D-LC MS and off-line 2D-LC MS. Multidimensional protein identification technology (MudPIT) is a common technique for on-line 2D-LC MS[18], where the column consists of SCX material back-to-back packed with reversed phase material inside a fused silica capillary[20]. RPLC is compatible with SCX due to its capability to remove salts. MudPIT operates in cycles where the salt concentration is increased

in each cycle to elute a portion of peptides out of the SCX stationary phase, followed by increasing percentage of an organic solvent (acetonitrile) to elute separated peptides into the ionization source and then to the tandem mass spectrometer.

Unlike on-line 2D-LC MS, the off-line 2D-LC MS first conducts peptide separation in a SCX LC which has a relatively large sample loading capacity and collects peptide fractions by a fraction collector of HPLC, then each individual fraction is injected into a RPLC MS/MS instrument for detection. In my thesis work, I applied an improved approach of off-line 2D-LC MS that was first described by Wang et al[21]. In this approach, an extra RPLC separation step is done prior to loading the RPLC-MS/MS instrument. The RPLC is equipped with a UV detector so the amount of peptides flushed through the RP column can be calculated based on their UV absorbance and a standard calibration curve. The collected peptide fractions are then concentrated to an optimal amount and are finally injected into the RPLC-MS/MS instrument. By controlling the sample amount injected, this approach allows optimal data collection by tandem MS.

1.2 MS instrumentation

1.2.1 Ionization techniques

In the past decades, MS-based proteomics research has experienced rapid growth due to important breakthroughs in experimental methods, instrumentation, and data processing approaches. One of the most important developments is the introduction of soft ionization method that enables detection of peptides or protein in MS because proteins and peptides are polar, non-volatile and thermally unstable molecules that require transfer into the gas phase without extensive degradation. Two ionization methods establish the basis of modern proteomics MS: matrix-assisted laser desorption ionization (MALDI)[22-24] and electrospray ionization (ESI)[25].

1.2.1.1 Matrix assisted laser desorption ionization (MALDI)

MALDI produces ions from a solid. MALDI requires a solid matrix that absorbs laser energy and transfer it to acidified analyte where the rapid laser heat results in desorption of matrix and release of $[M+H]^+$ ions of analyte into the gas phase. In order to achieve acceptable signal-to-noise ratio for detection of peptide ions, hundreds of laser shots are generally required[26]. Ions generated by MALDI are usually singly charged and therefore suitable for top-down analysis of high-molecular-weight proteins with pulse analysis instruments. The drawbacks of MALDI are low shot-to-shot reproducibility, high background signals in low m/z range and strong dependence on sample preparation methods[27,28].

1.2.1.2 Electrospray ionization (ESI)

In contrast to MALDI, ESI is a solution-based ionization technique. Electrospray ionization requires high voltage (2–6 kV) applied between the emitter at the end of the separation pipeline and the inlet of the mass spectrometer. Such high voltage generated electrically charged spray, Taylor cone[29], followed by formation and desolvation of analyte-solvent droplets (Figure 1.1). Formation and evaporation of the droplets is aided by either a heated capillary or by sheath gas flow at the mass spectrometer inlet. Several different physical models of ESI ion formation have been proposed[30,31]. However, the common practical features are the generation of multiply charged species, sensitive to the concentration of analyte and low flow rate. Micro and nano-ESI are an important development in ESI technique[32,33], where the flow rates are lowered to a nanoliter-per-minute range to enhance the sensitivity.

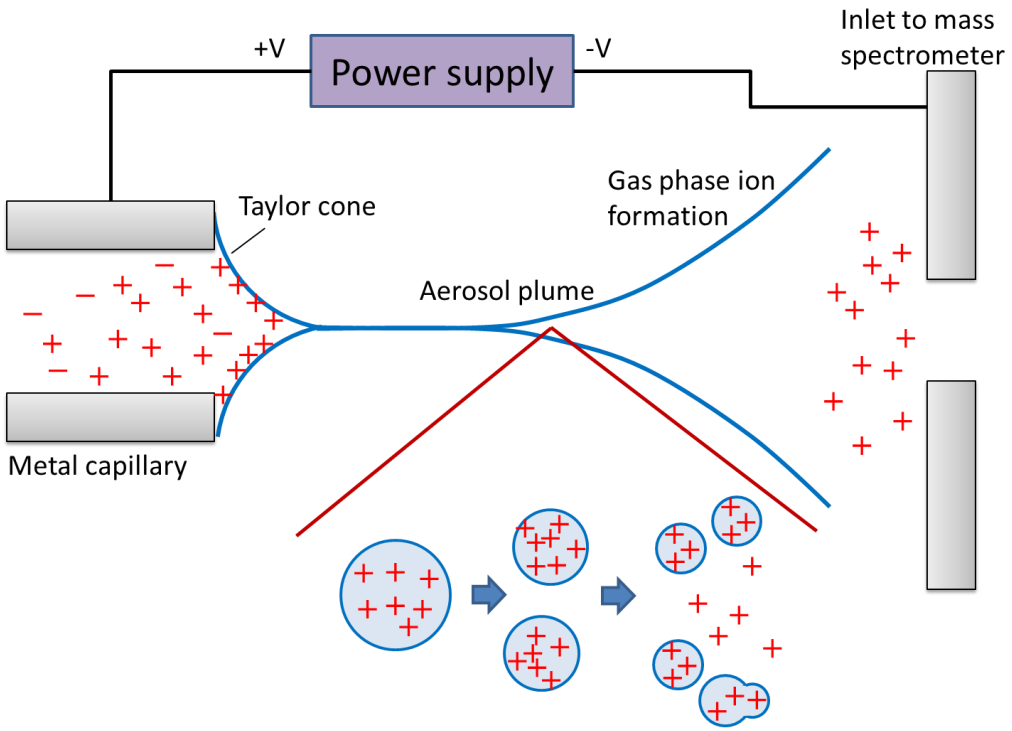


Figure 1.1 Process of electrospray ionization (ESI). (Adapted from Harris, 2007[34])

1.2.2 Mass analysers

Mass analyzers are a critical component of each MS instrument because they can store ions and have them separated based on the mass-to-charge ratios. Ion trap, Orbitrap, and ion cyclotron resonance (ICR) mass analyzers separate ions based on their m/z resonance frequency, quadrupoles (Q) utilize m/z stability, and time-of-flight (TOF) analyzers use flight time. Each mass analyzer has its own properties, such as mass range, analysis speed, resolution, sensitivity, ion transmission, and dynamic range. Hybrid mass spectrometers have been built that combine more than one mass analyzer for specific needs during analysis. In-depth introduction of each type of mass analyser is beyond the scope of this chapter. I will focus on introduction of two mass analysers used in my thesis work. They are quadrupole time-of-flight instrument (QTOF) for proteomics work and Fourier transform ion cyclotron resonance (FT-ICR) for metabolomics work.

1.2.2.1 QTOF-MS

QTOF combines a quadrupole mass analyzer and a time-of-flight tube to form a tandem mass spectrometer.

Figure 1.2 is a schematic demonstration of the QTOF Premier system from Waters which consists of an ESI source, a quadrupole unit, a collision cell and an orthogonal acceleration TOF tube. This hybrid orthogonal acceleration time-of-flight mass spectrometer provides automated accurate mass measurement of precursor and fragment ions (<30 ppm) to yield high confidence in structural elucidation and database search results[35]. Other features of QTOF Premier include the ZSpray source with high transmission efficiency and NanoLockSpray interface allows flow rate of 5-1000 $nlmin^{-1}$ for electrospray ionization.

The quadrupole contains four parallel metal rods in square or near-square configuration where each opposing rod pair is connected together electrically and a direct current (DC) and a radio frequency (RF) voltage is applied between one pair of rods and the other. When ions travel through the rods, only those with a certain mass-to-charge ratio (m/z) have stable trajectory path while the trajectory paths of other ions with different (m/z) are not stable and therefore cannot go through the quadrupole. In other words, the quadrupole is an ion filter that only allows ions with certain m/z ratio reach the TOF tube.

The TOF tube is a field free region; no electric or magnetic fields are applied across the flight tube. The ions are pulsed in by an electric field applied on the top of the flight tube. The velocity of an accelerated ion is given as:

$$v = \left(\frac{2qU}{m}\right)^{\frac{1}{2}} \qquad \text{(Equation 1.1)}$$

Where U is the voltage, q is the charge of the ion and m is the mass of the ion. The flight time in the TOF is determined by:

$$t = \frac{d}{v} \qquad \text{(Equation 1.2)}$$

Where d is the length of ion path. Since U and d are constant in a given flight tube with an electric field of known strength, the ion velocity or its flight time is determined by the ion's m/z ratio only.

A V-mode reflectron TOF is generally used to increase the mass analyzer's resolution. The reflectron can partially compensate for initial energy dispersion of ions and focus ions having the same m/z value to the detector by using an ion reflector. The ion reflector consists of successive sets of plates, within which an electric field gradient is created. When ions with different kinetic energy enter this field, higher energy ions will penetrate deeper into the reflectron, increasing their flight path length and observed flight time. Compared with a linear TOF analyzer, the reflectron TOF increases mass resolution ($m/\Delta m$ = ~10,000 to 20,000), with minimal losses in sensitivity.

There are three modes available in QTOF analysis, TOF MS, TOF MS/MS, and data directed analysis (DDA). In the TOF MS mode, resolving DC voltage is off and only RF is applied to the quadrupole, all the ions could go through quadrupole and be detected by TOF. In the TOF MS/MS mode, resolving DC voltage is applied to selectively allow one specific mass to get through quadruple or scan through a wide mass range to select candidate ions for fragmentation. (Precursor ion scan). DDA mode is just when the instrument is set to an automatic switch between TOF MS and TOF MS/MS.

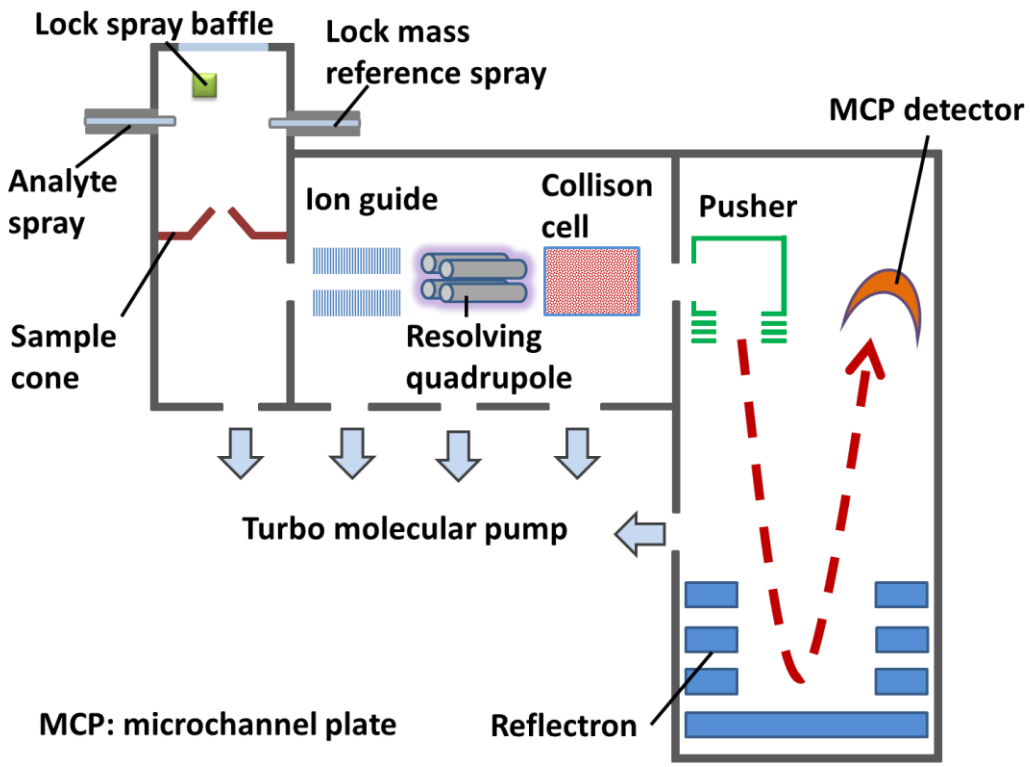


Figure 1.2 Schematic diagram of ESI QTOF MS from Waters

1.2.2.2 FT-ICR-MS

Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS), also known as Fourier transform mass spectrometry (FT-MS), is a type of mass analyzer (or mass spectrometer) for determining the mass-to-charge ratio (m/z) of ions based on the cyclotron frequency of the ions in a fixed magnetic field[36]. Because the magnetic field is generated by a super-conducting magnet and is much more stable than RF electric field, FT-ICR MS can provide ultimate limit of detection and precision of mass measurement to enable metabolomics analyses.

In the basic FT-MS instrument, the ions are generated in the source and then pass through a series of pumping stages at increasingly high vacuum. When the ions enter the cell, pressures are in the range of 10^{-10} to 10^{-11} mBar. The cell is located inside a spatial uniform static superconducting high field magnet (typically 4.7 to 13 Tesla; in the Bruker instrument used in my work, it is a 9-

Tesla instrument) cooled by liquid helium and liquid nitrogen. When the ions pass into the magnetic field they are bent into a circular motion in a plane perpendicular to the field (Figure 1.3) by the Lorentz Force (Equation 1.3).

$$\boldsymbol{F} = m\left(\frac{dv}{dt}\right) = qv \times \boldsymbol{B} \qquad \textbf{(}\text{Equation 1.3)}$$

Where **F** is the Lorentz Force observed by the ion when entering the magnetic field; **B** is the magnetic field strength (constant); *v* is the incident velocity of the ion; m is the ion mass; q is the charge on the ion.

If no collision occurs, the ions keep a constant speed in vacuum and the magnetic field bends the path of ions into circles with characteristic radius r determined by equation 1.4:

$$\boldsymbol{\omega} = \frac{q\boldsymbol{B}}{m} \qquad \text{(Equation 1.4)}$$

Where **ω** represents the cyclotron frequency. According to equation 1.4, the experimentally measured ion cyclotron frequency can be correlated to an ionic mass-to-charge ratio. The cyclotron frequency of an ion is inversely proportional to its mass-to-charge ratio (m/q) and directly proportional to the strength of the applied magnetic field. Ions with lower m/q have higher cyclotron frequencies

First introduced by Comisarow and Marshall[37], the ICR trap of modern FT-ICR-MS contains three pairs of plate electrodes which play different roles. As shown in Figure 1.4, trapping plates prevent ions from escaping out of the cell during the detection. Excitation plates provide a swept RF pulse field cross the ICR cell. Each individual excitation frequency will couple with the ions natural motion and excite them to a higher orbit where they induce an alternating imaging current between the detector plates.

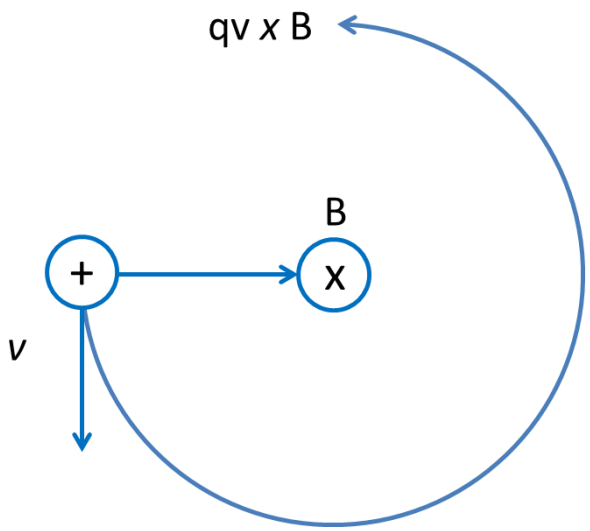


Figure 1.3 Ion cyclotron motion. Moving path of the positive ion in the plane is bent into a circle by the Lorentz magnetic force generated by a homogenous magnetic field perpendicular to the plane. (Adapted from Marshall et al., 1998[36])

There are four events sequentially occurs in a typical FT-ICR-MS experiment. First, a large negative voltage is applied to the trapping plates to remove all ions leftover in the cell from the previous experiment. Second, an electron beam or a laser beam is used to ionize molecules in the ICR cell or a packet of ions is introduced into the ICR cell. Then an RF sweep is applied to excitation plates to excite ions with all the m/z ratio to larger cyclotron orbits. Finally, image currents can be detected by detector plates and the signal is amplified and deconvoluted by fast Fourier transformation into frequency vs. intensity spectrum followed by converting the resulted spectrum into mass spectrum.

In FT-ICR-MS, ICR cell is naturally an ion trap, thus the measurement of ions is non-destructive. This feature enables MS/MS or even MS^n for unknown compound structural elucidation. In addition, excitation waveforms can be manipulated to excite ions with selected mass range during FT-ICR tandem mass spectrometry experiments.

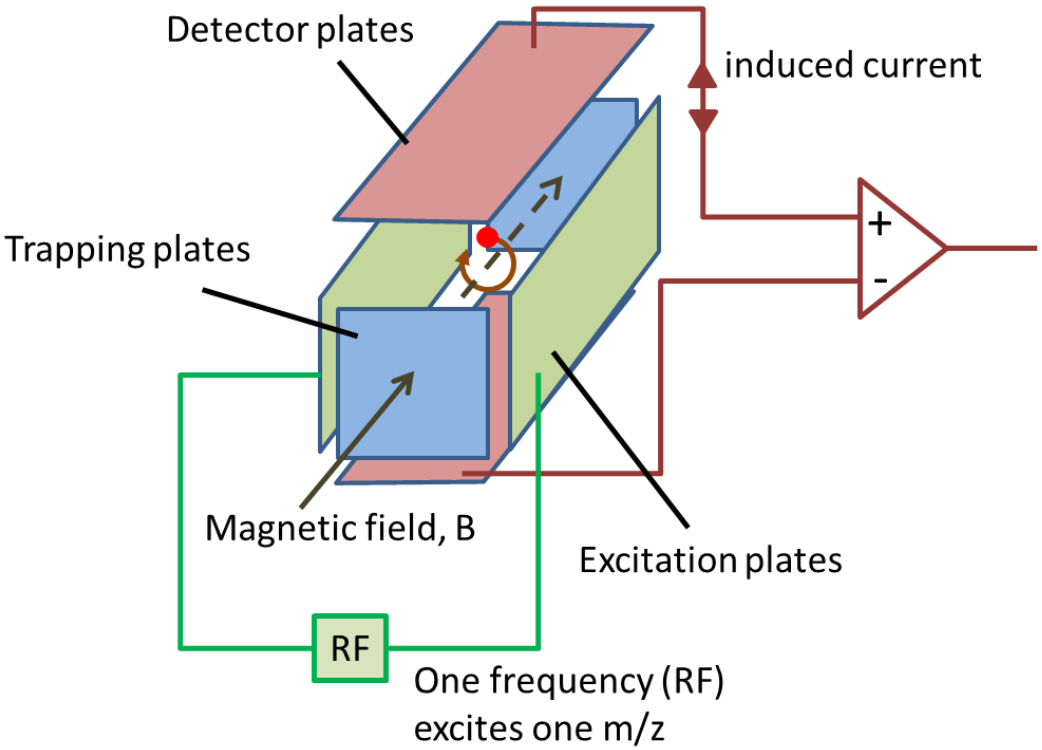


Figure 1.4 Schematic demonstration of an ICR cell. The plates positioned on the front and the back are two trapping plates, on the two sides are the excitation plates, and the two on the top and bottom are the detection plates. (Adapted from website http://www.chm.bris.ac.uk/ms/theory/fticr-massspec.html)

1.2.3 Tandem mass spectrometry (MS/MS)

Tandem mass spectrometry, also known as MS/MS or MS^2, contains multiple steps of mass spectrometry selection with certain type of fragmentation occurring in between the stages. MS/MS offers additional information about specific ion of interest and is therefore critical for identification of peptide sequences in MS-based proteomics. In this approach, ions of interest are selected by m/z from the first round of MS detection and are subjected to fragmentation by a number of dissociation methods. Among them, collision-induced dissociation (CID)[38] is one of the most popular methods for peptide ion fragmentation. In this method, the precursor ions are accelerated in vacuum by an electric field to high kinetic

energy. These energetic ions then collide with inert gas molecule (such as He, Ar, N_2) in the collision cell. Such collision converts part of the kinetic energy into internal energy of the precursor ions which results in fragmentation of precursor ions into smaller fragment ions. The small fragment ions are then detected by a mass spectrometer. For proteome analysis, the collision energy of CID is typically ranging from 10 – 100 eV in most mass spectrometers (triple quadruple, ion traps, QTOF, etc.)[38].

In the gas phase, several bonds in the peptide backbone can possibly be broken under CID (Figure 1.5A). At low energy CID, the fragment peptide ions are dominated by fragment ions resulted from cleavage of the amide bonds in the peptide backbone. The nomenclature differentiates fragment ions according to which end of the fragment retains a charge after fragmentation and where the bond breakage occurs[38]. As shown in Figure 1.5B, the fragment ion is named as b ion when the charge associated with the peptide ion retains on the N-terminus of the broken backbone. On the other hand, if the charge stays at the C-terminus of the broken backbone, this ion is then named as y ion. Every b or y ion contains a subscript which is used to designate the specific amide bond fragmented to generate the observed fragment ions: the subscript of b ions is the number of amino acid residues present on the fragment ion counted from the amino-terminus, whereas the subscript of y ions shows the number of amino acids counted from the carboxyl-terminus.

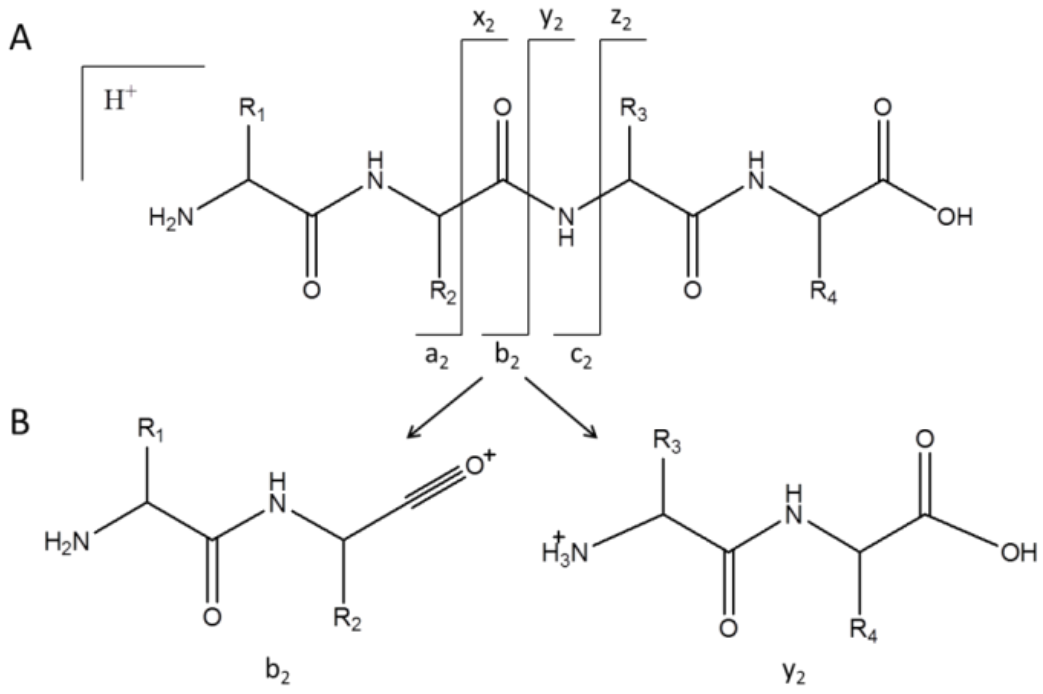


Figure 1.5 (A) CID fragmentation pattern of a peptide ion, (B) an example of b ions and y ions.

It is practically feasible to identify a protein by tandem MS analysis of their fragment peptides[39,40]. When acquiring MS/MS spectra from a tandem mass spectrometer, the first MS scan (or called survey scan) interrogates the mass/charge ratio of each peptide, performs on-the-fly data process and ranks the peptide ions based on their relative intensities. Four to eight most intense peptide ions will then be chosen for further MS/MS analysis, starting from the most intense peptide ion. The selected peptide ion is first guided to the collision cell and fragmented during the MS/MS scan. A MS/MS of the peptide ion is then acquired and serves as a unique "fingerprint" of the peptide. Upon completion of MS/MS spectral collection for those four to eight most intense peptide ions, the next survey scan is performed, followed by the MS/MS scans. This cycle continues till the LC separation of peptides is completed. Then all the acquired

MS/MS spectra from the LC-MS/MS run are processed and imported into a MS/MS database search engine for protein identification.

There are a few database searching engines available for interpreting MS/MS spectra and identifying peptide sequences[14,41-43]. The most popular database searching engine includes: SEQUEST (http://fields.scripps.edu/sequest/), MASCOT (http://www.matrixscience.com/) and X!Tandem (http://www.thegpm.org/TANDEM/). Basically, experimental fragment ion MS/MS spectra are matched against theoretical fragment ion spectra for all the peptides in the databases that have the same precursor ion mass within the user-defined experimental error. Peptides that turn out to be the first hit along with the identification scores equal or higher than the identity threshold defined by the database are generally considered as positive matches. The identification of proteins is then achieved by correlating the matched peptides to their corresponding proteins according to the sequence matching.

1.3 MS-based quantitative proteomic analysis

The large-scale proteomics for systems biology has a key benefit which is the capability to quantify functional entities of the cell, i.e., the proteins. The objective of such measurements is to acquire concentrations of proteins associated with different states of the cell. Quantitative measurement of protein concentrations is one of the key steps toward constructing a functional network. There are two broad groups of quantitative methods in MS-based proteomics: (a) relative quantitative proteomics and (b) absolute quantitative proteomics. Relative quantitative proteomics can compare two or more samples by getting concentration ratios of individual proteins. It could provide information of differentially expressed proteins in different states. On the other hand, if internal standards with known concentrations are used, the absolute concentration of proteins can be determined by comparing the ratio of the peak intensities between

sample proteins and internal standards. In the following sections, the strategies of these two approaches will be discussed.

1.3.1 Relative quantitative proteomics

1.3.1.1 Label-free method

When isotopic labeling method is not applicable, label-free methods can be used for estimation of protein abundance in proteomics. Label-free techniques use either spectral counting or peptide ion intensity for quantifying protein abundance[44] which will be discussed in the following sections.

Spectral counting exploits the frequency of MS/MS sequencing event as a surrogate for estimating protein abundance. It has been reported that the frequency of MS/MS sequencing event is directly proportional to the abundance of proteins[45]. Comparison of spectral counting method with $^{14}N/^{15}N$ labeling method shows strong correlations between these two techniques for quantification in MudPIT[46]. The result of spectral counting is usually presented as the number of peptides detected per protein but it fails to consider the fact that larger proteins and proteins with many peptides in the preferred mass range for mass spectrometry will generate more observed peptides. The introduction of exponentially modified protein abundance index (emPAI) has minimized this problem and increased the accuracy of spectral counting results[47]. The main advantage of spectral counting is its universal applicability but its accuracy is protein-dependent and thus can only be considered as a semi-quantitative strategy for protein abundance estimation[48].

In contrast to spectral counting methods, the ion intensity method performs protein abundance analysis by comparing the ion current intensities in extracted ion chromatograms of the same targeted peptide from different samples[49]. Briefly, samples are prepared using the same protocol and run in LC-MS under an identical condition. One of the samples is chosen as control and its MS/MS spectra are used for peptide identification and then the ion peaks of each peptide

serves as reference to extract and reconstruct the extracted ion chromatograms of the corresponding peptides in other samples. Finally, the total area of ion peaks of the same peptide is integrated through the whole chromatogram for each sample and the relative values are reported. Although the direct comparison of peptide signal across multiple runs is straightforward, in reality feature alignment and correction is not practically trivial and is still an active field of study. In addition, the result of ion intensity method could be greatly affected by the variation of LC-MS/MS instrumental performance which results in reduced accuracy of peptide matching[50,51].

To sum up, compared to label-based methods (see below), label-free methods do not require the labeling step and have broader applicability. However, they are less accurate and can only be regarded as semi-quantitative approaches for estimating protein abundance.

1.3.1.2 Label-based method

Compared to label-free methods, label-based method is an approach for relative quantitation that is more expensive, labor-intensive and time-consuming but less sensitive to experimental bias than label-free methods. The label-based methods involve labeling the samples with stable isotope labels that could be distinguished by the mass spectrometer between identical proteins in separate samples. For instance, isotopic tags are one type of label that consists of stable isotopes incorporated into protein functional group that cause a known mass shift of the labeled protein or peptide in the mass spectrum. Differentially labeled samples are mixed and analyzed together in MS, and the differences in the peak intensities of the isotope pairs accurately reflect relative difference in the concentration of the corresponding proteins. Isotopic tags can be introduced by (a) metabolically (b) chemically and enzymatically[52], as described below.

Metabolic labeling. The proteome in cells can be metabolically labeled with isotopic tags by growing cells on culture medium containing the stable isotopes of

elements (^{15}N) or stable isotopes of amino acids (heavy Arg, Lys, Leu, and Ile). For example, total labeling of yeast was achieved using ^{15}N-enriched cell culture media[53]. The relative ratio of peptides is obtained by comparing heavy/light peptide pairs in MS, and then protein levels are derived from statistical evaluation of the peptide ratios. Metabolically labeled samples could be combined at the early stage of sample preparation which in turn reduces experiment variation for measurement. One major limitation of metabolic labeling is the requirement of culturing samples in a special medium with isotopic nutrient; most of clinical samples could not match such a requirement.

The most widely used metabolic labeling method is the stable isotope labeling by amino acids in cell culture (SILAC) approach where cell media contain $^{13}C_6$-Lys and $^{13}C_6$, and $^{15}N_4$-Arg for comprehensive labeling of tryptic cleavage products[52]. Stable isotope labeling by amino acids in cell culture (SILAC) was dated back in 2002[54]. Several applications of SILAC in *in vivo* metabolic ^{15}N labeling of model organisms such as *Caenorhabditis elegans*, *Drosophila melanogaster*[55], and rat[56] has been reported. SILAC has been proved as a powerful method for studies in cell signaling, post translation modifications such as phosphorylation[57,58], protein–protein interaction[59] and regulation of gene expression[60]. In addition, SILAC has become an important approach in the global study of secreted proteins and secretory pathways[61].

Chemically and enzymatically labeling. Chemical derivatization procedures can be applied to any sample at either the protein or the peptide level with advantages of equal reactivity and identical performance in downstream sample preparation. There are two large classes of chemical labeling approaches: non-isobaric and isobaric tags. Non-isobaric labeling methods acquire quantitation information from MS spectra, while isobaric labeling methods obtain quantitation data from MS/MS spectra. Peptides reacted with non-isobaric tags will have certain mass shift in an MS spectrum depending on the number of labels. One typical example of non-isobaric tags is dimethylation on peptide or protein free

amine groups with isotopic coded formaldehyde, ^{12}C-/^{13}C- or H/D. The labeled peptides are 2 to 6 Da mass shifted in MS spectra. Our group has developed a related technique called N-terminal dimethylation after lysine guandination (2-MEGA) which could improve the accuracy of quantitation[62] and facilitate data analysis with an introduction of a fixed 6-Da mass shift. Another example of non-isobaric labeling is isotope-coded affinity tags (ICAT), used for the labeling of free cysteine[63]. Example of isobaric tags is isobaric tags for relative and absolute quantification (iTRAQ), used for the labeling of free amines[64].

In addition to chemical derivatization, post-biosynthetic labeling of proteins and peptides can be done by enzymatic derivatization in vitro. Enzymatic labeling usually incorporates ^{18}O either during or after digestion using trypsin[65,66]. Originally, this technique has been applied to aid *de novo* sequencing of peptides by mass spectrometry[67], but recently has caught attention in quantitative proteomic applications[68]. Basically, samples are separately digested with trypsin in $H_2{}^{18}O$. During the digestion process, trypsin substitutes the ^{16}O atoms at the C-terminus of peptides with ^{18}O. Incorporation of ^{18}O into C-termini of peptides results in a mass shift of 2 Da per ^{18}O atom and trypsin introduce 2 oxygen atoms resulting in 4 Da shift in mass which is generally sufficient for differentiation. One disadvantage of this labeling technique is that full labeling is difficult to achieve and peptides with different labeling rates complicate data analysis[69,70].

1.3.2 Absolute quantitative proteomics

There has been a long history of using isotope labeling method in quantitative mass spectrometry[71]. It has now become more commonly used techniques called AQUA (absolute quantification of proteins)[72]. In the simplest case, a known quantity of a stable labeled protein standard is added to a protein mixture, followed by trypsin digestion, and the absolution quantification can be achieved by comparing the mass spectrometric signal of the isotope peptide to the endogenous peptide in the sample. Unlike the metabolic labeling, the relative quantification is acquired for a large number of proteins in the mixture, and the

addition of protein standards to a proteome sample focuses on the determination of the quantity of one or a few specific proteins of interest. This approach is very useful for studies such as the analysis and validation of biomarkers in a large number of clinical samples[73] or measuring the level of particular peptide post-translational modification[74,75].

The approach usually utilize synthetic genes to express concentrated protein standards which upon tryptic digestion provides multiple peptides of the same protein for quantification or as quantification standards for a group of proteins.

Provided that tryptic digestion of proteome sample results in very complex mixture of peptides and most mass spectrometers have limited dynamic range for detection, the AQUA approach has some limitations. One disadvantage is that it is very ambiguous to determine how much of the standard should be added to a sample. This quantity may be very distinct among all the proteins of interest and their expression levels may vary greatly in a sample. Another limitation is that the specificity of the spiked standards may be compromised due to the possible presence of multiple isobaric peptides in the mixture. Both of the problems could be addressed by multiple reaction monitoring (MRM) where the mass spectrometer monitors the intact peptide mass as well as one or more particular fragment ions from the same proteins over the course of LC-MS experiment. This technique practically eliminates ambiguities in peptide assignments and enhances the quantification dynamic range to 4-5 orders of magnitudes[76]. Despite the ability to determine protein quantities from an AQUA experiment, there are still concerns regarding the accuracies of results because any sample manipulation prior to adding the standards may introduce bias toward the results (either losses or enrichment). Therefore, the amount of a protein determined by AQUA experiment might not correctly reflect the real expression levels of this protein in a sample[77].

1.4 MS based quantitative metabolomics analysis

Metabolomics is the comprehensive study of all endogenous metabolites in a biological system. The term metabolome was first coined by Oliver in 1998[4] and Fiehn first defined metabolomics in 2001 as the comprehensive and quantitative analysis of all metabolites that could help in the understanding of biosystems and revealing of their metabolome[78]. Metabolites participate in all metabolic reaction and are end products of cellular processes. Their concentration can be regarded as the definitive response of a biosystem to generic and environmental influences[79,80]. Therefore the detection, identification and quantification of metabolites are significant for the better understanding of the organism. Although metabolomics is not as mature a field as other omics technologies, its combination with genomics, transcriptomics, and proteomics offers possibility of insightful biological studies and can also be used in disease biomarker discovery[81,82].

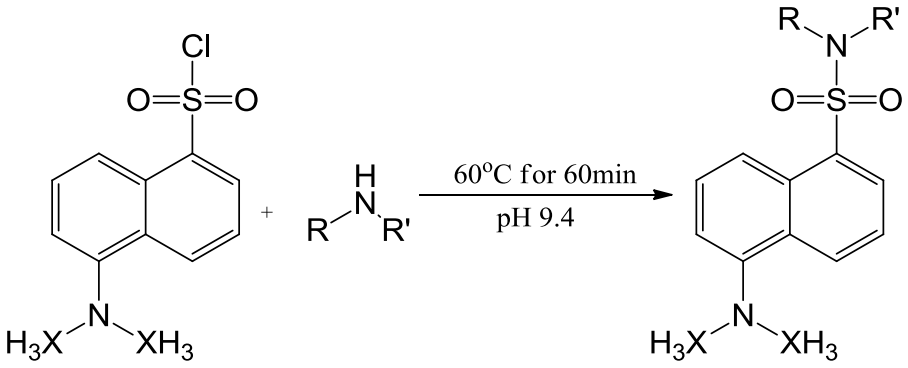
The metabolomics study was initially performed using GC-MS[83] and also relies on nuclear magnetic resonance (NMR), while LC-MS has increasingly been used nowadays. One approach to achieve comprehensive metabolome coverage is to combine different analytical platforms such as NMR, GC-MS, and LC-MS and compile their results together, taking the advantages of all the analytical techniques. With this approach, Wishart et al. successfully profiled 308 metabolites in human cerebrospinal fluid (CSF)[84].

An ideal metabolomics study should provide a comprehensive, qualitative and quantitative overview of all metabolites in a biological system. However, the diversity of their chemophysical properties, large dynamic ranges, and sheer number of metabolites hindered the comprehensive detection of all metabolites. Currently, the strategy to fractionate the metabolome into several groups according to their hydrophobicity, chemical structures or other properties is frequently used. Different group of metabolites are then analyzed by different

optimized LC-MS methods focusing on the targeted group of metabolites and all the results are compiled together in the end.

1.4.1 Dansylation labeling

For metabolome analysis, it is essential to profile all the metabolites qualitatively and quantitatively. One of the many challenges include the detection of the highly polar and hydrophilic metabolites, because they are poorly retained on the RPLC column and elute at the column initial void volume where ESI-MS detection sensitivity is reduced due to poor ESI desolvation performance in high water mobile phase. Moreover, the co-elution of polar species and salts will worsen the ion suppression effect. Different methods were applied to overcome this problem. Hydrophilic interaction chromatography (HILIC) was used to separate polar and hydrophilic compounds, but suffers from low separation efficiency. Chemical derivatization is more often used as it not only improves chromatographic behavior, but also increases MS detectability.



X = ^{12}C or ^{13}C

Figure 1.6 Reaction scheme for labeling amine and phenol containing metabolites using isotope coded dansyl chloride (light chain, x = 12; heavy chain, x = 13)

Dansylation is frequently used as derivatization method for quantification of amino acids, biogenic amines and phenol containing metabolites. It is simple,

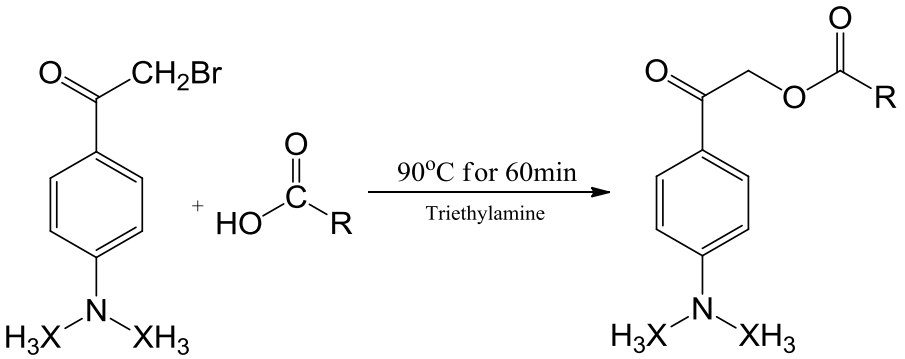
robust, and readily combined with HPLC separation and fluorescence or UV detection[85,86]. It is also used to form derivatives of targeted analytes, which can be detected by LC-MS[87,88].

With dansylation (Figure 1.6), a relative hydrophobic naphthalene moiety and a more easily protonated dimethylamino moiety is introduced to the metabolites containing primary amine, secondary amine or phenolic hydroxyl group(s). The labeled metabolites would have better chromatographic properties and be eluted out at higher percentage of organic solvent. This will ensure a better ionization desolvation efficiency and electrospray stability. Plus the hydrophobic nature of naphthalene moiety can increase droplet affinity of the analyte, the overall ESI response is dramatically increased. In addition, the easily ionized dimethylamino moiety on the dansyl group makes the competition of charged metabolites to the droplet surface more favorable. Finally, the m/z of labeled metabolites is shifted to the higher mass region and the stability of the metabolites is improved, which will all result in better LC-MS performance.

1.4.2 Carboxylic acid labeling

It is important to achieve global profiling of metabolites with carboxyl groups in metabolomics because a large portion of the metabolome contains these groups including a vast number of fatty acids and TCA cycle acids. For instance, about 65 % of the ~5000 known endogenous human metabolites contain at least one carboxylic acid group in a chemical structure[89]. Derivatization of carboxylic acids can be done using commercially available phenacyl bromide (PBr) and it is originally used for improving performance of UV detection[90]. Inspired by this technique, our group developed a new isotope labeling method, based on the use of isotope-coded *p*-dimethylaminophenacyl (DmPA) bromide as a reagent, combined with liquid chromatography-mass spectrometry (LC-MS) for high performance metabolome analysis with a focus on profiling carboxylic acid-containing metabolites[8]. Compared to PBr, the use of DmPA allows the introduction of a mass tag and concurrent improvement in LC-MS analysis.

Figure 1.7 shows the structure of the reagent, DmPA bromide, and the reaction scheme for labeling the carboxylic acid to form isotope mass-coded derivatives. The mass difference of the ^{13}C-/^{12}C-labeled products with one tag has a nominal mass of 2 Da.



X = ^{12}C or ^{13}C

Figure 1.7 Reaction scheme for labeling carboxylic acid-containing metabolites using isotope coded *p*-dimethylaminophenacyl (DmPA) bromide (light chain, x = 12; heavy chain, x = 13).

There are several advantages of using DmPA derivatization for analysis of acid metabolites. First, DmPA derivatization enhances the hydrophobicity of labeled acids and improves the separation of labeled metabolites by RPLC while many unlabeled acids cannot retained on a RP column. Second, DmPA derivatization enhances the ESI efficiency significantly where several factors contribute to this signal enhancement. The increased propensity of being charged for the labeled acid due to the presence of the dimethylamine moiety attached to the aromatic ring of the tag where a tertiary amine can be readily formed. The labeled acid has higher hydrophobicity making it easier to stay on the surface of the droplets during ESI. An elution solvent with higher organic solvent content where a labeled acid is eluted out compared to the unlabeled counterpart. The third advantage is that a proper isotope mass tag of DmPA can be readily attached to a carboxylic acid containing metabolite, and the labeled metabolite does not

display any isotope effect on RPLC. Finally, DmPA isotope labeling facilitates the identification of metabolite peaks with less interference in LC-MS[8].

To conclude, with dansylation and acid labeling approaches, larger numbers of metabolites in each of the sub-metabolome could be detected, resulting in a better coverage of amine, phenol and carboxylic acid containing sub-metabolomes.

1.4.3 Quantitative differential isotope labeling

For the quantitative metabolomic analysis using LC-MS, peak intensity cannot accurately represent metabolites abundance as ESI efficiency is affected by ion suppression from sample matrix or other co-eluting compounds. Stable-isotope-labeled (SIL) analogues are commonly used as internal standards to overcome matrix interference and ion suppression effects, but are limited to the analysis of small number metabolites. It is not practical to get SIL analogues of every metabolite in the metabolome as SIL analogues are sometimes not affordable, hard to synthesize or even the metabolite itself is with an unknown identity.

Differential-isotope-labeling (DIL) method, however, uses one chemical labeling reaction to introduce an isotope tag to a sub-group of analytes of one sample and another mass-different isotope tag to these analytes in the comparative sample. After individual sample labeling, the differential-isotope-labeled samples are mixed for LC-MS analysis and the intensity ratio between isotope-labeled peak pairs provides the basis for metabolites quantification. For relative quantification of two comparative samples, a light isotope-tag and a heavy isotope-tag need to be introduced to the two comparative samples. As for absolute quantification, the comparative sample needs to be a set of standards with known concentration. DIL is widely used in quantitative proteomic analysis[63,91,92] but have fewer applications in quantitative metabolomics analysis. Early reports of DIL for metabolite analysis used iTRAQ reagent, which is commonly known as a labeling reagent for quantitative proteomic analysis to label amino acids for

quantitative analysis of small molecules in urine and blood samples. Guo et al. from our research group has also reported a series of $^{12}C/^{13}C$ isotope labeling methods for the analyses of amine-, phenol- and carboxyl-containing metabolites[7,8,93]. Other quantitative metabolomic analysis include chemical derivatization of analytes with isotope-coded reagents followed by GC-MS, NMR analysis[94,95] or cell culturing using isotope enriched media[96].

Figure 1.8 shows the experimental workflow of $^{12}C/^{13}C$ isotope labeling strategy used in my work to compare metabolomic changes of SG1 grown in the presence and absence of butanol. One advantage of this strategy is that the $^{12}C/^{13}C$ isotope labeled metabolites are coeluted and detected by MS and there is no isotopic effect on the RPLC separation and ionization. Moreover, the differential isotope label has been incorporated in a tag, and thus all of the metabolites containing a specific functional group can be labeled in a single reaction. In other words, all the targeted internal standards can be generated in one step.

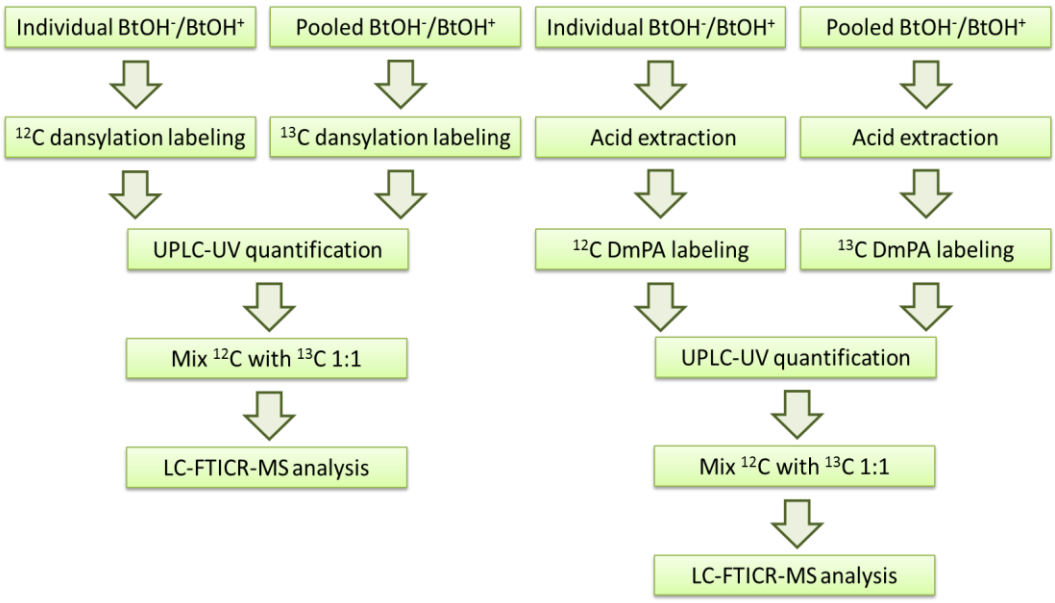


Figure 1.8 Experimental work flow of quantitative metabolomic analysis of SG1 in the presence and absence of butanol.

1.5 Scope of the thesis work

My thesis work focuses on using mass spectrometry based method to study a butanol tolerant strain of *Staphylococcus warneri* called SG1. In Chapter 2, I will describe the work of comparative proteome profiling of SG1 grown in the presence and absence of 1.5 % 1-Butanol using 2D-LC-MS/MS. In Chapter 3, I will describe the quantitative proteomic analysis using 2-MEGA isotope labeling combined with 2D-LC-MS/MS and quantitative metabolomic analysis of SG1 in the presence and absence of butanol using targeted isotope labeling with LC-FTICR-MS. $^{12}C/^{13}C$ dansylation labeling was performed to quantify amine and phenol containing metabolites and $^{12}C/^{13}C$ DmPA labeling was performed for the quantification of carboxylic acid containing metabolites. The biological significances of the proteome and metabolome results are discussed. In Chapter 4, I will conclude my thesis work and discuss some future work related to my research.

1.6 Literature cited

(1) Yates, J. R.; Osterman, A. L. *Chemical Reviews* **2007**, *107*, 3363.

(2) Wilkins, M. R. *Biotechnology* **1996**, *14*, 61.

(3) Tyers, M.; Mann, M. *Nature* **2003**, *422*, 193.

(4) Oliver, S. G.; Winson, M. K.; Kell, D. B.; Baganz, F. *Trends in biotechnology* **1998**, *16*, 373.

(5) Theodoridis, G.; Gika, H. G.; Wilson, I. D. *TrAC Trends in Analytical Chemistry* **2008**, *27*, 251.

(6) Dunn, W. B.; Ellis, D. I. *TrAC Trends in Analytical Chemistry* **2005**, *24*, 285.

(7) Guo, K.; Li, L. *Analytical Chemistry* **2009**, *81*, 3919.

(8) Guo, K.; Li, L. *Analytical Chemistry* **2010**, *82*, 8789.

(9) McDonald, W. H.; Yates, J. R., 3rd. *Disease Markers* **2002**, *18*, 99.

(10) McDonald, W. H.; Yates, J. R., 3rd *Current Opinion in Molecular Therapeutics* **2003**, *5*, 302.

(11) Kelleher, N. L. *Analytical Chemistry* **2004**, *76*, 196 A.

(12) Yates, J. R., 3rd *Annual Review of Biophysics and Biomolecular Structure* **2004**, *33*, 297.

(13) Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R. *Nature Biotechnology* **1999**, *17*, 676.

(14) Eng, J. K.; McCormack, A. L.; Yates Iii, J. R. *Journal of the American Society for Mass Spectrometry* **1994**, *5*, 976.

(15) Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *20*, 3551.

(16) Yates, J. R.; Ruse, C. I.; Nakorchevsky, A. *Annual Review of Biomedical Engineering* **2009**, *11*, 49.

(17) Simpson, D. M.; Beynon, R. J. *Journal of Proteome Research* **2010**, *9*, 444.

(18) Fournier, M. L.; Gilmore, J. M.; Martin-Brown, S. A.; Washburn, M. P. *Chemical Reviews* **2007**, *107*, 3654.

(19) Tomas, R.; Kleparnik, K.; Foret, F. *Journal of Separation Science* **2008**, *31*, 1964.

(20)Wolters, D. A.; Washburn, M. P.; Yates, J. R., 3rd *Analytical Chemistry* **2001**, *73*, 5683.

(21) Wang, N.; Xie, C.; Young, J. B.; Li, L. *Analytical Chemistry* **2009**, *81*, 1049.

(22) Hillenkamp, F.; Karas, M. *Methods in Enzymology* **1990**, *193*, 280.

(23) Hillenkamp, F.; Karas, M.; Beavis, R. C.; Chait, B. T. *Analytical Chemistry* **1991**, *63*, A1193.

(24) Karas, M.; Hillenkamp, F. *Analytical Chemistry* **1988**, *60*, 2299.

(25) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64.

(26) Liao, P.-C.; Allison, J. *Journal of Mass Spectrometry* **1995**, *30*, 763.

(27) Silvertand, L. H.; Torano, J. S.; de Jong, G. J.; van Bennekom, W. P. *Electrophoresis* **2008**, *29*, 1985.

(28) Zheng, J.; Li, N.; Ridyard, M.; Dai, H.; Robbins, S. M.; Li, L. *Journal of Proteome Research* **2005**, *4*, 1709.

(29) Taylor, G. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **1964**, *280*, 383.
(30) Felitsyn, N.; Peschke, M.; Kebarle, P. *International Journal of Mass Spectrometry* **2002**, *219*, 39.
(31) Thomson, B. A.; Iribarne, J. V. *The Journal of Chemical Physics* **1979**, *71*, 4451.
(32) Griffin, P. R.; Coffman, J. A.; Hood, L. E.; Yates Iii, J. R. *International Journal of Mass Spectrometry and Ion Processes* **1991**, *111*, 131.
(33) Emmett, M. R.; Caprioli, R. M. *Journal of the American Society for Mass Spectrometry* **1994**, *5*, 605.
(34) Harris, D. C. *Quantitative Chemical Analysis, 7th edition*; W. H. Freeman, 2007.
(35) Bristow, T.; Constantine, J.; Harrison, M.; Cavoit, F. *Rapid Communications in Mass Spectrometry* **2008**, *22*, 1213.
(36) Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. *Mass Spectrometry Reviews* **1998**, *17*, 1.
(37) Comisarow, M. B. *International Journal of Mass Spectrometry and Ion Physics* **1981**, *37*, 251.
(38) Wysocki, V. H.; Resing, K. A.; Zhang, Q.; Cheng, G. *Methods* **2005**, *35*, 211.
(39) Sadygov, R. G.; Cociorva, D.; Yates, J. R., 3rd *Nature Methods* **2004**, *1*, 195.
(40) Delahunty, C.; Yates, J. R., 3rd *Methods* **2005**, *35*, 248.
(41) Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *20*, 3551.
(42) Nesvizhskii, A. I. *Journal of Proteomics* **2010**, *73*, 2092.
(43)Craig, R.; Cortens, J. P.; Beavis, R. C. *Journal of Proteome Research* **2004**, *3*, 1234.
(44) Domon, B.; Aebersold, R. *Science* **2006**, *312*, 212.
(45) Liu, H.; Sadygov, R. G.; Yates, J. R., 3rd *Analytical Chemistry* **2004**, *76*, 4193.

(46) Zybailov, B.; Coleman, M. K.; Florens, L.; Washburn, M. P. *Analytical Chemistry* **2005**, *77*, 6218.
(47) Ishihama, Y.; Oda, Y.; Tabata, T.; Sato, T.; Nagasu, T.; Rappsilber, J.; Mann, M. *Molecular & Cellular Proteomics* **2005**, *4*, 1265.
(48) Xia, Q.; Hendrickson, E. L.; Wang, T.; Lamont, R. J.; Leigh, J. A.; Hackett, M. *Proteomics* **2007**, *7*, 2904.
(49) Ryu, S.; Gallis, B.; Goo, Y. A.; Shaffer, S. A.; Radulovic, D.; Goodlett, D. R. *Cancer Informatics* **2008**, *6*, 0.
(50) Paulovich, A. G.; Billheimer, D.; Ham, A.-J. L.; Vega-Montoto, L.; Rudnick, P. A.; Tabb, D. L.; Wang, P.; Blackman, R. K.; Bunk, D. M.; Cardasis, H. L.; Clauser, K. R.; Kinsinger, C. R.; Schilling, B.; Tegeler, T. J.; Variyath, A. M.; Wang, M.; Whiteaker, J. R.; Zimmerman, L. J.; Fenyo, D.; Carr, S. A.; Fisher, S. J.; Gibson, B. W.; Mesri, M.; Neubert, T. A.; Regnier, F. E.; Rodriguez, H.; Spiegelman, C.; Stein, S. E.; Tempst, P.; Liebler, D. C. *Molecular & Cellular Proteomics* **2010**, *9*, 242.
(51) Tabb, D. L.; Vega-Montoto, L.; Rudnick, P. A.; Variyath, A. M.; Ham, A.-J. L.; Bunk, D. M.; Kilpatrick, L. E.; Billheimer, D. D.; Blackman, R. K.; Cardasis, H. L.; Carr, S. A.; Clauser, K. R.; Jaffe, J. D.; Kowalski, K. A.; Neubert, T. A.; Regnier, F. E.; Schilling, B.; Tegeler, T. J.; Wang, M.; Wang, P.; Whiteaker, J. R.; Zimmerman, L. J.; Fisher, S. J.; Gibson, B. W.; Kinsinger, C. R.; Mesri, M.; Rodriguez, H.; Stein, S. E.; Tempst, P.; Paulovich, A. G.; Liebler, D. C.; Spiegelman, C. *Journal of Proteome Research* **2009**, *9*, 761.
(52) Ong, S.-E.; Mann, M. *Nature Chemical Biology* **2005**, *1*, 252.
(53) Oda, Y.; Huang, K.; Cross, F. R.; Cowburn, D.; Chait, B. T. *Proceedings of the National Academy of Sciences of the United States of America* **1999**, *96*, 6591.
(54) Jiang, H.; English, A. M. *Journal of Proteome Research* **2002**, *1*, 345.
(55) Krijgsveld, J.; Ketting, R. F.; Mahmoudi, T.; Johansen, J.; Artal-Sanz, M.; Verrijzer, C. P.; Plasterk, R. H.; Heck, A. J. *Nature Biotechnology* **2003**, *21*, 927.
(56) Wu, C. C.; MacCoss, M. J.; Howell, K. E.; Matthews, D. E.; Yates, J. R., 3rd *Analytical Chemistry* **2004**, *76*, 4951.

(57) Ibarrola, N.; Kalume, D. E.; Gronborg, M.; Iwahori, A.; Pandey, A. *Analytical Chemistry* **2003**, *75*, 6043.

(58) Ibarrola, N.; Molina, H.; Iwahori, A.; Pandey, A. *Journal of Biological Chemistry* **2004**, *279*, 15805.

(59) Selbach, M.; Mann, M. *Nature Methods* **2006**, *3*, 981.

(60)Schwanhäusser, B.; Gossen, M.; Dittmar, G.; Selbach, M. *Proteomics* **2009**, *9*, 205.

(61) Hathout, Y. *Expert Review of Proteomics* **2007**, *4*, 239.

(62) Wang, P.; Lo, A.; Young, J. B.; Song, J. H.; Lai, R.; Kneteman, N. M.; Hao, C.; Li, L. *Journal of Proteome Research* **2009**, *8*, 3403.

(63) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. *Nature Biotechnology* **1999**, *17*, 994.

(64) Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlet-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J. *Molecular & Cellular Proteomics* **2004**, *3*, 1154.

(65) Yao, X.; Freas, A.; Ramirez, J.; Demirev, P. A.; Fenselau, C. *Analytical Chemistry* **2001**, *73*, 2836.

(66)Reynolds, K. J.; Yao, X.; Fenselau, C. *Journal of Proteome Research* **2002**, *1*, 27.

(67) Rose, K.; Simona, M. G.; Offord, R. E.; Prior, C. P.; Otto, B.; Thatcher, D. R. *The Biochemical Journal* **1983**, *215*, 273.

(68) Miyagi, M.; Rao, K. C. *Mass Spectrometry Reviews* **2007**, *26*, 121.

(69) Johnson, K. L.; Muddiman, D. C. *Journal of the American Society for Mass Spectrometry* **2004**, *15*, 437.

(70) Ramos-Fernandez, A.; Lopez-Ferrer, D.; Vazquez, J. *Molecular & Cellular Proteomics* **2007**, *6*, 1274.

(71) Desiderio, D. M.; Kai, M. *Biomedical Mass Spectrometry* **1983**, *10*, 471.

(72) Gerber, S. A.; Rush, J.; Stemman, O.; Kirschner, M. W.; Gygi, S. P. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100*, 6940.
(73) Pan, S.; Zhang, H.; Rush, J.; Eng, J.; Zhang, N.; Patterson, D.; Comb, M. J.; Aebersold, R. *Molecular & Cellular Proteomics* **2005**, *4*, 182.
(74) Kirkpatrick, D. S.; Denison, C.; Gygi, S. P. *Nature Cell Biology* **2005**, *7*, 750.
(75) Kirkpatrick, D. S.; Gerber, S. A.; Gygi, S. P. *Methods* **2005**, *35*, 265.
(76) Wolf-Yadlin, A.; Hautaniemi, S.; Lauffenburger, D. A.; White, F. M. *Proceedings of the National Academy of Sciences of the United States of America* **2007**, *104*, 5860.
(77) Bantscheff, M.; Schirle, M.; Sweetman, G.; Rick, J.; Kuster, B. *Analytical and Bioanalytical Chemistry* **2007**, *389*, 1017.
(78) Fiehn, O. *Comparative and Functional Genomics* **2001**, *2*, 155.
(79) Fiehn, O. *Plant Molecular Biology* **2002**, *48*, 155.
(80) Rochfort, S. J.; Towerzey, L.; Carroll, A.; King, G.; Michael, A.; Pierens, G.; Rali, T.; Redburn, J.; Whitmore, J.; Quinn, R. J. *Journal of Natural Products* **2005**, *68*, 1080.
(81) Li, X.; Yang, S.; Qiu, Y.; Zhao, T.; Chen, T.; Su, M.; Chu, L.; Lv, A.; Liu, P.; Jia, W. *Metabolomics* **2010**, *6*, 109.
(82) Wei, J.; Xie, G.; Zhou, Z.; Shi, P.; Qiu, Y.; Zheng, X.; Chen, T.; Su, M.; Zhao, A.; Jia, W. *International Journal of Cancer* **2011**, *129*, 2207.
(83) Dalgliesh, C. E.; Horning, E. C.; Horning, M. G.; Knox, K. L.; Yarger, K. *The Biochemical Journal* **1966**, *101*, 792.
(84) Wishart, D. S.; Lewis, M. J.; Morrissey, J. A.; Flegel, M. D.; Jeroncic, K.; Xiong, Y.; Cheng, D.; Eisner, R.; Gautam, B.; Tzur, D.; Sawhney, S.; Bamforth, F.; Greiner, R.; Li, L. *Journal of chromatography. B* **2008**, *871*, 164.
(85) Loukou, Z.; Zotou, A. *Journal of Chromatography A* **2003**, *996*, 103.
(86) Minocha, R.; Long, S. *Tree Physiology* **2004**, *24*, 55.
(87) Xia, Y. Q.; Chang, S. W.; Patel, S.; Bakhtiar, R.; Karanam, B.; Evans, D. C. *Rapid Communications in Mass Spectrometry* **2004**, *18*, 1621.

(88) Cech, N. B.; Enke, C. G. *Mass Spectrometry Reviews* **2001**, *20*, 362.
(89) 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.; Shaykhutdinov, R.; Li, L.; Vogel, H. J.; Forsythe, I. *Nucleic Acids Research* **2009**, *37*, D603.
(90) Durst, H. D.; Milano, M.; Kikta, E. J.; Connelly, S. A.; Grushka, E. *Analytical Chemistry* **1975**, *47*, 1797.
(91) Ong, S. E.; Foster, L. J.; Mann, M. *Methods* **2003**, *29*, 124.
(92) Karlsson, K. E.; Novotny, M. *Analytical Chemistry* **1988**, *60*, 1662.
(93) Guo, K.; Ji, C.; Li, L. *Analytical Chemistry* **2007**, *79*, 8631.
(94) Ye, T.; Mo, H.; Shanaiah, N.; Gowda, G. A.; Zhang, S.; Raftery, D. *Analytical Chemistry* **2009**, *81*, 4882.
(95) Wade, D. *Chemico-Biological Interactions* **1999**, *117*, 191.
(96) Bennett, B. D.; Yuan, J.; Kimball, E. H.; Rabinowitz, J. D. *Nature Protocols* **2008**, *3*, 1299.

Chapter 2. Comprehensive Proteomic Profiling of *Staphylococcus warneri* SG1 Cultured in the Presence and Absence of Butanol using Shotgun Proteomics

2.1 Introduction

Solvent tolerant microorganisms are often exploited for biofuel production and bioremediation applications. In both cases, the bacterium must survive in an environment containing a high titer of organic solvent that would normally be toxic to non-adapted cells. These organic compounds are detrimental to the cells due to the chaotropic effects on cell membranes and disruption of various biological processes such as the respiration chain, nutrient transport, signal transduction, etc[1-4]. Both Gram-positive and Gram-negative bacteria can be intrinsically tolerant to organic solvents but the latter are studied more rigorously because of their additional outer membrane and thus are generally more tolerant[5]. Despite the numerous studies undertaken to investigate solvent tolerance, the complicated mechanism still remains elusive due to its wide scope and sophistication.

Staphylococcus warneri is a solvent tolerant Gram-positive bacterium that constitutes a part of the human skin flora. A recently isolated *Staphylococcus warneri* strain (SG1) carries tolerance to alkanes, short-chain alcohols, and cyclic aromatic compounds[6]. In particular, SG1 could grow in the presence of 2.5 % 1-butanol, making it an excellent candidate chassis for biofuel production. The genome of SG1 has been sequenced[6], representing the "blue-print of life" but no proteomic analysis has been conducted. The genome consists of 2.56 Mbases and is estimated to encode 2457 open reading frames (ORFs). Proteomic analysis cannot be replaced by any genome or transcriptome analysis as proteins perform biological functions and make up cellular structures, their post translational modifications are not coded by genes, and there is poor correlation between gene

*Mingguo Xu contributed to the data processing of this work (^{15}N spectra validation part). Victor Cheng contributed partially to the data interpretation of this work.

activity and protein abundance[7-9]. Up to now, proteomic analysis has not reached the same scale as genomic analysis. In most proteomics applications, only a fraction of the proteome is examined. This disparity is mainly due to the complexity of the proteome and the difficulty of characterizing proteins (such as membrane proteins), compared to genome and DNA analysis. The use of bacterial proteome information to address important biological questions relies heavily on the development of new and improved proteomic technologies.

Mass spectrometry (MS) studies, in combination with two-dimensional gel electrophoresis (2-DE), was used by Mao et al. to probe proteomic differences between a wild type *Clostridium acetobutylicum* strain (DSM1731) and a mutant strain (Rh8) that has higher butanol tolerance and yield[10,11]. In total, 564 and 341 proteins were identified in the intracellular and membrane fractions respectively, which combine to represent a mere 23.5 % of the predicted proteome. Their work illustrates the power of the proteomic approach for discovering biologically significant proteins involved in butanol tolerance and production. However, the relatively small number of identified proteins highlights the difficulty in characterizing the proteome of bacteria using gel-based method. Susanne et al. used 2D-LC analysis of the cytosolic proteome and 1D gel-LC studies of membrane proteins and have identified another 473 proteins in addition to the 745 proteins identified by 2D-PAGE, leading to a total of 1218 proteins identified in exponentially growing *B. subtilis* cells[12]. Otto et al. also reported a GeLC-MS approach that uses 1D SDS-PAGE coupled with LC-MS/MS analysis for proteome profiling of *Bacillus subtilis*, leading to 52 % coverage of the predicted open reading frames (ORFs)[13]. Unell et al. employed a bottom-up 2D-LC-MS/MS approach for proteome profiling of *Arthrobacter chlorophenolicus* in different growth conditions and identified 3749 proteins that covered over 70 % of the predicted proteome[14].

The overall goal of this study was to obtain a comprehensive whole cell proteome profiling data so that we can increase our knowledge on the physiology

of this newly isolated laboratory strain of *Staphylococcus warneri*. We applied a two-dimensional liquid chromatography (2D-LC) MS shotgun method with optimal precursor ion extraction (PIE) strategy and maximal LC-MS sample loading technique on SG1 grown in the presence and absence of butanol. Spectra validation using metabolic isotope labeling was applied to validate the spectrum-to-sequence assignment, thereby increasing the confidence of identification and in order to construct a more reliable MS/MS spectral library. To understand the biological mechanism of butanol tolerance of this species, the emPAI approach was used to roughly quantify proteins and compare differences in protein abundance in SG1 grown with or without 1-Butanol in the media.

2.2 Experimental Procedures

2.2.1 Chemicals and reagents.

All the chemicals and reagents, unless otherwise stated, were purchased from Sigma-Aldrich Canada (Markham, ON, Canada). Lysostaphin was purchased from AMBI Products (Lawrence, NY).Phosphoric acid (H_3PO_4), potassium chloride (KCl), potassium dihydrogen phosphate (KH_2PO_4), and ammonium bicarbonate (NH_4HCO_3) were purchased from EMD Chemical, Inc. (Mississauga, ON, Canada). Sequencing grade modified trypsin, HPLC grade formic acid, LC-MS grade water, acetone, and acetonitrile (ACN) were purchased from Fisher Scientific Canada (Edmonton, AB, Canada). A domestic 900 W (2450 MHz) sunbeam microwave oven was used to perform microwave-assisted protein solubilization experiments.

2.2.2 Cell growth and protein sample preparation.

Staphylococcus warneri strain SG1 were grown in Luria Bertani broth at 37 ^{o}C with shaking for 16 hours. For MS studies, 2 L cultures were grown in triplicate (seeded with a 0.1 % inoculum), with or without 1.5 % (V/V) 1-Butanol, and harvested by centrifugation at 8,000 x *g* for 15 minutes and resuspended in

100 mM Tris / 5 mM EDTA buffer, pH 7.0. Cell lysis was carried out either mechanically by repeated passage (4x) through a Constant Systems TS benchtop cell disruptor (Daventry, Northants., UK) at 40 kpsi, or enzymatically by adding NaCl (100 mM), lysostaphin (10 μgmL^{-1}) and lysozyme (50 μgmL^{-1}), followed by incubation at 37 ^{o}C for 1 hour. Unbroken cells and cell debris were removed by centrifugation at 10,000 x *g* for 20 minutes, and the cell lysates were frozen immediately with liquid nitrogen.

Protein assays were performed to adjust protein concentrations of lysates to similar levels. Acetone, pre-cooled to -80 ^{o}C, was gradually added to the whole cell lysates to a final concentration of 80 % (V/V) and the mixtures were incubated overnight at -80 ^{o}C. Samples were then spun at 20,800 ×*g* for 20 minutes and the pellets were washed with 40 µL of pre-chilled acetone before drying at room temperature. The pellets were then subjected to microwave-assisted protein solublization in urea[15]. Briefly, 8 M urea was added to the whole cell lysates followed by six cycles of microwave irradiation in 30 s cycles with sample cooling and homogenization between cycles. The mixtures were then centrifuged at 20,800 ×*g* for 5 minutes and the pellets were subjected to a fresh round of microwave assisted urea extraction. Upon complete solubilization of the pellets, the supernatant fractions were pooled and diluted with 100 mM NH_4HCO_3 to reduce the urea concentration to ~1 M. Samples were analyzed by protein assay and reduced with dithiothreitol for 1 h at 37 ^{o}C, followed by alkylation with iodoacetamide for 0.5 h at room temperature in the dark. Trypsin was added to a protein:tryspin ratio of 40:1 and incubated at 37 ^{o}C for 20 h for complete digestion. The tryptic digests were acidified with 50 % trifluoroacetic acid to pH 2 and injected into an Agilent 1100 HPLC system (Palo Alto, CA) for desalting and quantification. A Polaris C18-A column (4.6 mm × 50 mm,3 µm particle size, 300 Å pore size) (Varian,Palo Alto, CA) was used for desalting and a UV detector operating at 214 nm was used for quantification of the eluted peptides[16].

2.2.3 Proteome profiling by 2D-LC-MS/MS

After protein digestion, the desalted peptides were dried, reconstituted in 0.2 % H_3PO_4 (pH 2.0) and separated by SCX liquid chromatography using a polysulfoethyl A column (2.1 mm x 250 mm, 5 μm particle size, 300 Å pore size) (PolyLC, Columbia, MD). Peptides were separated into 20 fractions using the following gradient: mobile phases A (10 mM KH_2PO_4, pH 2.76) and B (10 mM KH_2PO_4, pH 2.76, 500 mM KCl); t = 0 min, 0 % B; t = 1 min, 4 % B; t = 17 min, 20 % B; t = 39 min, 60 % B; t = 45 min, 100 % B; t = 50 min, 100 % B; t = 52 min, 0 % B; t = 62 min, 0 % B. The collected peptide fractions were then desalted and quantified. Less abundant neighbor fractions were combined to a total of 20 fractions.

The SCX fractionated peptides were further separated by reversed phase liquid chromatography (RPLC) using a nano ACQUITY Ultra Performance LC system (Waters, Missisauga, ON) with an Atlantis dC18 column (75 μm × 150 mm, 3μm particle size, 100 Å pore size) (Waters, Milford, MA).The following gradient was applied to separate the peptides: mobile phases A (0.1 % FA in water) and B (0.1 % FA in ACN); t = 0 min, 2 % B; t = 2 min, 7 % B; t = 85 min, 20 % B; t = 105 min, 30 % B; t = 110 min, 45 % B; t = 120 min, 90 % B; t = 125 min, 90 % B; t = 130 min, 95 % B[16]. The eluted peptides were then electrosprayed into an electrospray ionization (ESI) quadrupole time-of-flight (Q-TOF) Premier mass spectrometer (Waters, Missisauga, ON) at a flow rate of 350 $nLmin^{-1}$. A survey MS scan was acquired from m/z 350-1600 for 0.8 s, followed by 8 data-dependent MS/MS scans. A mass tolerance window of 80 mDa was applied for both dynamic and precursor ion exclusion[17], with a retention time tolerance window of 150 s. The collision energy used for MS/MS analysis was varied based on the precursor ion mass and charge state. A mixture of leucine enkephalin and (Glu1)-fibrinopeptide B, used as mass calibrants (i.e., lock-mass), was infused at a flow rate of 300 $nLmin^{-1}$, and a 1 s MS scan was acquired every 1 min throughout the run. For establishing spectra library, each SCX fraction was analyzed twice on the

RPLC-MS with a precursor ion exclusion list involved in each of the second run to eliminate redundant identification. For other profiling analysis using biological triplicate samples, each fraction was analysed once on the RPLC-MS.

Raw RPLC-MS data were lock-mass corrected, de-isotoped, and converted to peak list files with retention time information using the ProteinLynx Global Server 2.3.0. The peak list files were processed using MASCOT (Matrix Science, London, U.K. version 2.2.1) to attain peptide sequence information. Database search was restricted to the 2457 predicted ORFs from SG1. The search parameters were selected as follows: enzyme, trypsin; missed cleavages, 1; peptide tolerance, 0.2 Da; MS/MS tolerance, 30 ppm; peptide charge, (+1, +2, and +3); fixed modification, carbamidomethylation(C); variable modifications, oxidation (M), and carbamylation (N-term). Identified peptides with scores larger than the MASCOT threshold score at a 95 % confidence level were retained to generate the final $BtOH^-$ and $BtOH^+$ proteomes.

The false positive peptide matching rate of our analysis was gauged by the target-decoy searching strategy[18]. Briefly, the MS/MS spectra were searched against the forward proteome sequence and the reversed proteome sequence (decoy). The decoy peptide matches with scores above the threshold scores at the 95 % confidence level were then compared to the forward sequence matches. If the score of a MS/MS spectrum matched with a decoy peptide was higher than that of the same spectrum matched with a normal peptide, a false positive match was registered. The FDR was calculated by equation 2.1:

$$\text{FDR} = \frac{\text{false positives}}{(\text{false positives} + \text{true positives})} \qquad \text{(Equation 2.1)}$$

Molecular weight (Mw) and isoelectric point (pI) were predicted using the "Compute pI/Mw tool" from Expasy (http://web.expasy.org/compute_pi/) to characterize the SG1 proteome *in silico*.

Semi-quantitative analysis of identified proteins was determined by the exponentially modified protein abundance index (emPAI) based on protein coverage by the peptide matches in a database search result. emPAI values obtained from the MASCOT server were normalized using equation 2.2[19].

$$\text{Normalized emPAI} = \frac{\text{emPAI}_{\text{protein}}}{\text{emPAI}_{\text{total}}} * 100\ \% \qquad \text{(Equation 2.2)}$$

2.2.4 Spectra validation using ^{15}N-labeled SG1 ($BtOH^-$).

The SG1 cells were grown in ^{15}N enriched media (CGM-1000-N-S, Cambridge Isotope Laboratories Inc., MA) for 24 hours. After cell harvest and lysis, whole cell lysates were analyzed in the same way as unlabeled $BtOH^-$ with the exception of the MASCOT search wherein the isotopic mass of nitrogen was set to 15.0001. The search results, including original spectra, peptide sequences, ion score, MASCOT threshold score for identity, retention time for identified peptides, experimental molecular mass, calculated molecular mass of the peptide, difference (error) between the experimental and calculated masses and corresponding protein information were exported to Excel using in-house software. MS/MS spectra of unlabeled $BtOH^-$ were validated experimentally by ^{15}N-labeled $BtOH^-$ spectra as depicted in Figure 2.1[20]. Briefly, the intensity of all the unlabeled and ^{15}N-labeled spectra was normalized. Then, spectra from unlabeled and ^{15}N-labeled sequences with MASCOT scores higher than 13 were overlaid. The similarity score of the fragmentation patterns was calculated by equation 2.3 where L_i and U_i are the relative intensity of the same fragment ion, i, in the labeled spectrum and the unlabeled spectrum, respectively.

$$\text{dot product} = \frac{\sum(U_i L_i)}{\sqrt{\sum U_i^2 \sum L_i^2}} \qquad \text{(Equation 2.3)}$$

If the number of common fragment ions were larger than five and the similarity score of the fragmentation patterns (calculated by equation 2.3) were larger than 0.95, the unlabeled spectra will be considered validated. Replicate

spectra of same sequence were consolidated using weighted averaging to construct a consensus spectrum. Both Individual spectrum from single spectrum identification and consensus spectrum from redundant identification were processed to exclude noise. The validated unlabelled spectra were compiled to form the spectra library.

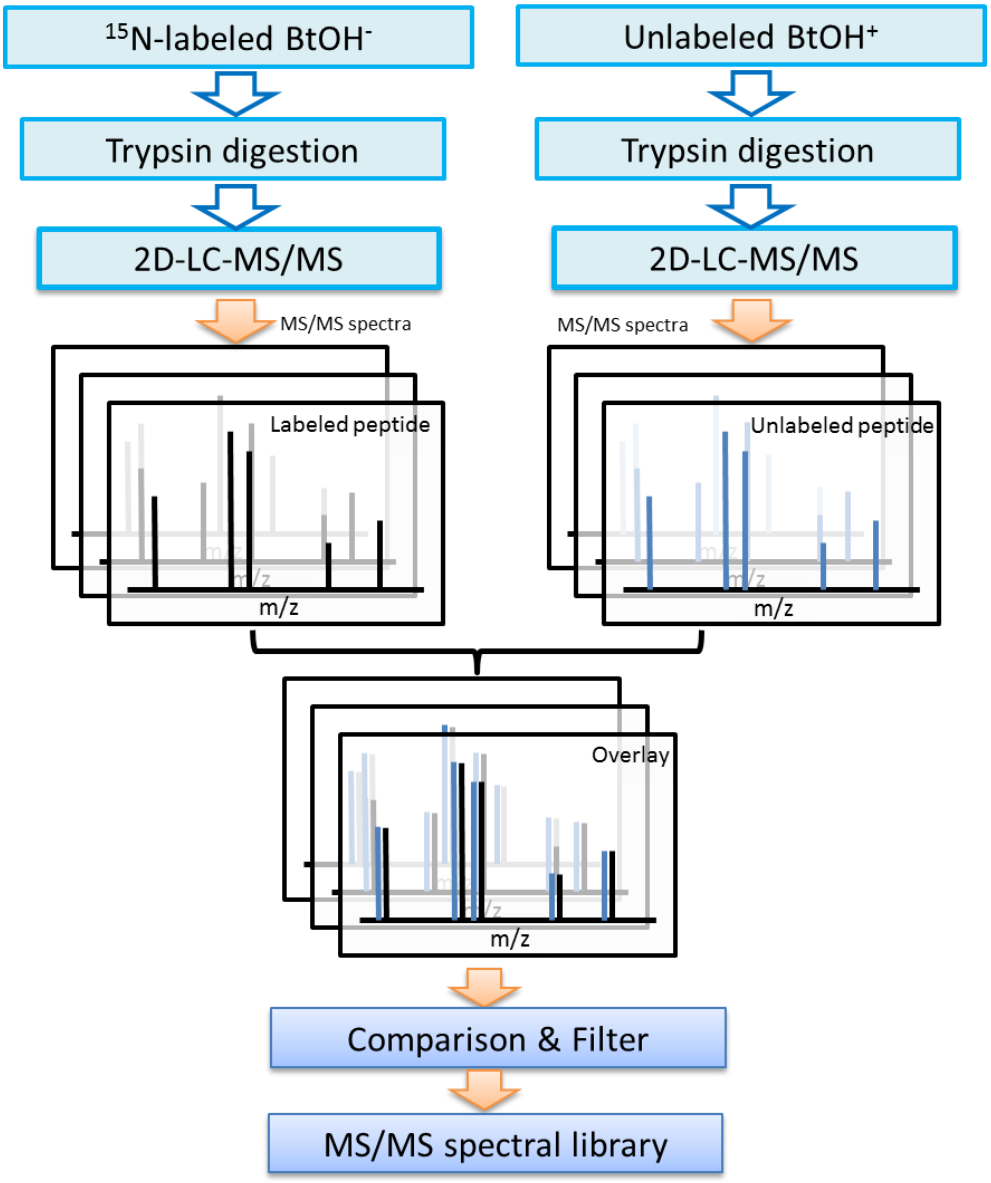


Figure 2.1 Experimental workflow of spectra validation using unlabeled and ^{15}N-labeled SG1 grown in BtOH$^-$ medium.

2.3 Results and Discussion

The complete genome sequence of a solvent tolerant Gram-positive bacterium, *Staphylococcus warneri* strain SG1, was recently published[6], but no proteomic analysis has been performed. This bacterium can thrive in the presence of short-chain alcohols, alkanes, esters and cyclic aromatic compounds (Table 2.1). In order to investigate the molecular mechanisms activated or repressed upon butanol challenge, we adopted a mass spectrometry (MS) approach to study the proteome of SG1 grown in the absence ($BtOH^-$) and presence ($BtOH^+$) of 1.5 % butanol, a concentration which decreased cell yield at stationary phase by approximately 15 %. An advantage of this study is that *Staphylococcus warneri* SG1 boasts a relatively compact proteome comprised of 2,457 proteins, which is considerably smaller than those of *Pseduomonas putida*[21,22] (around 5520 proteins) and *C. acetobutylicum*[10,11,23] (around 3850 protein coding genes) examined in similar studies.

Table 2.1 Growth of *Staph. warneri* SG1 in the presence of organic solvents. "++" "+" represent excellent and poor growth, respectively; "--" represents no growth. N.D.: not determined. Note that for organic solvents with high log P_{ow} values, the % values are only indicative of how much solvent was added, not the actual % due to its saturation in water.

Solvent	log P_{ow}	Growth on Solid Agar	Growth in Liquid media				
			0.5 %	1 %	2 %	4 %	6 %
Isobutanol	0.8	-	++	++	++	--	--
1-Butanol	0.8	-	++	++	+	--	--
Isoamyl alcohol	1.4	-	++	++	--	N.D.	N.D.
Chloroform	2.0	-	--	--	--	--	--
Benzene	2.0	+	++	--	--	--	--
1-Hexanol	2.0	-	--	--	--	--	--
Butyl butyrate	2.4	+	++	++	++	++	++
Toluene	2.5	+	++	++	++	++	--
1-Octanol	2.9	-	--	--	--	N.D.	N.D.
Hexane	3.5	+	++	++	++	++	++
1-Decanol	4.6	-	--	--	--	N.D.	N.D.
Octane	5.2	+	++	++	++	++	++
Decane	5.6	+	++	++	++	++	++

2.3.1 Spectra validation using ^{15}N-labeled SG1 ($BtOH^-$)

In order to increase protein identification confidence, we applied the ^{15}N isotopic labeling spectra validation strategy developed earlier in our group[20]. Table 2.2 shows the identification summary of the unlabeled $BtOH^-$ and ^{15}N-labeled $BtOH^-$ proteomes. A total of 130477 and 145009 MS/MS spectra were collected from unlabeled $BtOH^-$ and ^{15}N-labeled $BtOH^-$ samples, respectively. All of the MS/MS spectra were individually searched against SG1 database using MASCOT and generated two lists of peptide sequence matches. 19078 and 14910 spectra from unlabeled $BtOH^-$ and ^{15}N-labeled $BtOH^-$ samples were matched to peptide sequences with MASCOT scores of larger than 13, respectively. By overlaying the spectra of unlabeled and labeled matches of the same identification, the number of common fragment ions and the similarity score of intensity patterns of common ions can be readily calculated. After applying two filters of more than five common fragment ions and a similarity score of higher than 0.96 to the 57,679 comparisons, invalidated identifications were removed and 47,199 spectra pairs remained. In order to construct consensus spectrum for each peptide sequence assignment, replicated-spectra consolidation and noise reduction were performed. Finally, we were able to construct an MS/MS spectra library of tryptic digest for $BtOH^-$ sample with 3653 unique sequence and charge states corresponding to 3209 unique peptide sequences. Among the validated peptide sequences, 43 single-peptide matched or single-hit proteins were validated, 76 new peptides that were not initially identified at the 95% confidence level using database search alone were further identified.

There are several reasons why some identified peptides failed in the validation process. One possibility is that some peptides were only observed in the unlabeled spectrum and missing the ^{15}N-labeled counterpart, some ^{15}N-labeled spectra had significantly poorer quality, some unlabeled and ^{15}N-labeled pairs did not share enough common y- and b-ions or differed much in their fragmentation pattern. In the future, we could apply the optimal precursor ion inclusion strategy

to direct the spectra collection of same peptides in ^{15}N-labeled samples and arrange longer time on generating MS/MS spectra of the same peptides identified in unlabeled samples with the hope to increase the number of matched pairs and thus the number of validated spectra. Nevertheless, this work indicates that the database search strategy and parameters as described in the experimental section provides high confidence identification of the peptides. Specifically, when we compiled and researched these spectra using MASCOT, the estimated FDR decreased dramatically from 1.37 % to 0 %. As FDR is a statistical estimation of the matching quality of the whole data set, the value 0 % represents a very high matching quality and high identification confidence. In order to keep the same FDR for the original experimental data set without performing metabolic labeling validation, the MASCOT threshold needs to increase to 44. While such a high threshold filter can reject most of the incorrect PSMs (peptide-spectrum matches), it also has the potential to exclude many correct peptide matches, resulting in reduced sensitivity. Additionally, compared to only applying a FDR filter, the spectra validation experiment provides experimental support to validate peptide identifications.

Table 2.2 Summary of ^{15}N spectra validation result.

	Number
Total comparison	57679
Validated peptides	3653
Validated proteins	894
Validated single hit proteins	43
Validated peptides that were not identified with 95% CL	76

2.3.2 Profiling by 2D-LC-MS/MS in biological triplicates

Independent experiments were performed on biological triplicate samples of SG1 grown in $BtOH^-$ and $BtOH^+$ media. Protein identification results from biological triplicate experiments were compared and summarized in Table 2.3. More than 60% of proteins were detected in all biological triplicate experiments

and approximately 5-10% of proteins were identified only once (Figure 2.2). Combining data of SG1 cultured in $BtOH^-$ and $BtOH^+$ resulted in a pool of 1567 protein identifications. 189 and 201 unique proteins were detected in cells grown in $BtOH^-$ and $BtOH^+$ media, respectively, while 1177 common proteins were observed (Figure 2.3). Molecular weight (Mw) and isoelectric point (pI) were predicted using the "Compute pI/Mw tool" from Expasy (http://web.expasy.org/compute_pi/) to characterize the SG1 proteome *in silico*. The distribution of molecular weight (Figure 2.4) and isoelectric point (Figure 2.5) showed a very similar pattern in both $BtOH^-$ and $BtOH^+$ proteomes. The majority of identified proteins are located in the molecular weight range of 10 kDa to 100 kDa, whereas less than 10 % of detected proteins had molecular weights smaller than 10 kDa or larger than 100 kDa. The pI distributions in the $BtOH^-$ and $BtOH^+$ proteomes (Figure 2.5) show a classical bimodal distribution that is observed in other bacteria proteomes[24,25].

Table 2.3 Protein identification summary of SG1 grown in $BtOH^-$ and $BtOH^+$ media. Data were collected from biological triplicate experiments. All protein identification was based on 95% confidence level.

	1_$BtOH^-$	2_$BtOH^-$	3_$BtOH^-$	1_$BtOH^+$	2_$BtOH^+$	3_$BtOH^+$
Total spectra	65086	54108	67421	76456	63203	72818
Redundant peptides	7447	7672	10076	10143	7412	8998
Unique peptides	4477	4251	5547	5896	4169	4921
Unique proteins	1046	1022	1202	1183	1072	1126
False discovery rate (%)	1.50	1.60	1.40	1.44	1.83	1.92

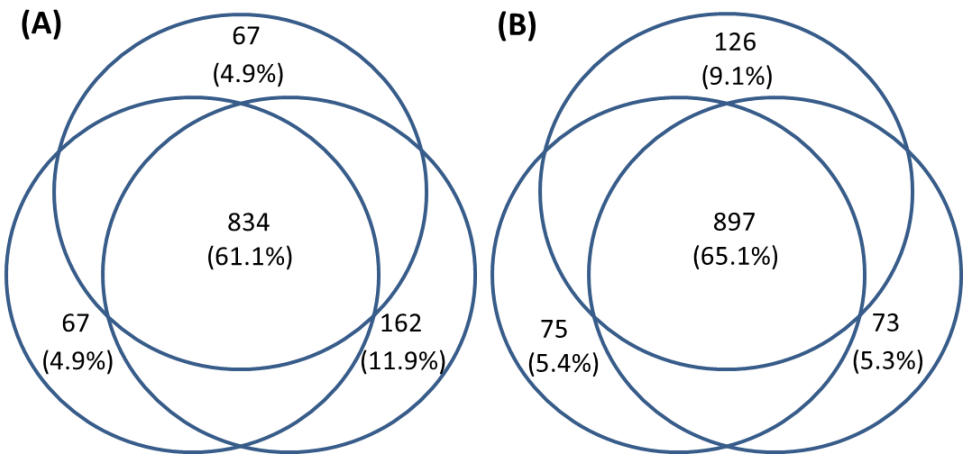


Figure 2.2 Venn diagrams showing protein identification comparison of the biological triplicate experiments of SG1 grown in $BtOH^-$ (A) and $BtOH^+$ media (B). Protein identification was carried out using 95 % confident level.

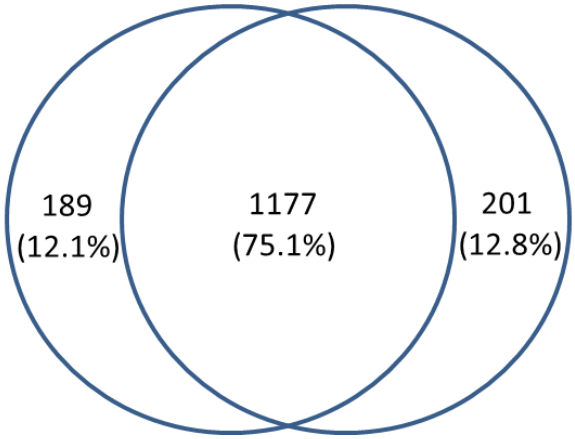


Figure 2.3 Venn diagram showing protein identification overlap of *Staph. warneri* SG1 grown in $BtOH^-$ (left) and $BtOH^+$ (right). Number of identified proteins was from the merged results from biological triplicate experiments.

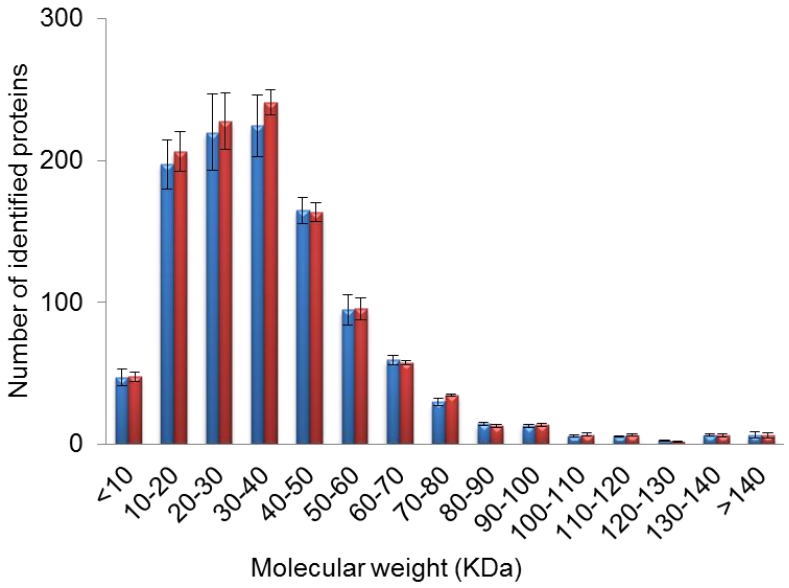


Figure 2.4 Molecular weight distribution of observed proteins in *Staph. warneri* SG1 grown in $BtOH^-$ (blue) and $BtOH^+$(red). Data were gathered from three independent experiments on biological triplicate samples.

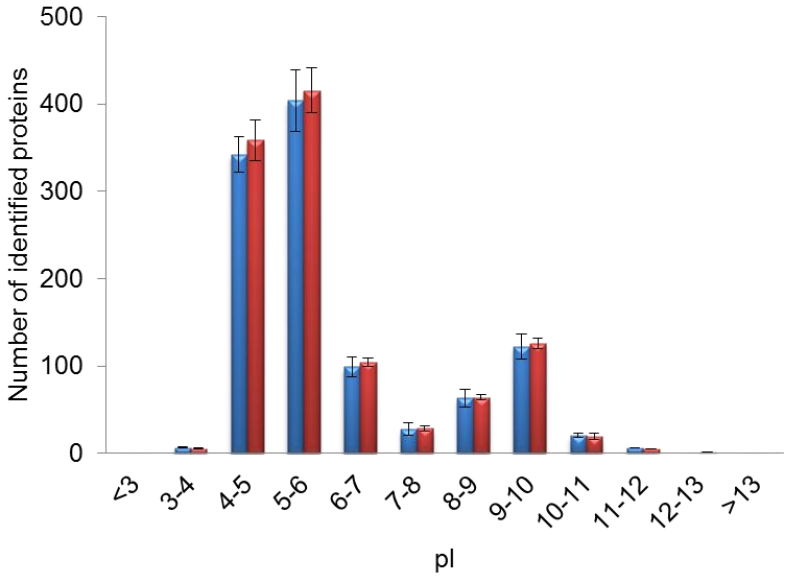


Figure 2.5 pI value distribution of identified proteins in *Staph. warneri* SG1 grown in $BtOH^-$(blue) and $BtOH^+$(red). Data were gathered from three independent experiments on biological triplicate samples.

Protein expression profiles were first analyzed by cluster of orthologous groups (COG)[26]. The observed proteomes of SG1 were grouped into 20 functional categories, as shown in Figure 2.6. Comparison of the $BtOH^-$ and $BtOH^+$ proteomes shows that their COG distribution profiles are highly similar which is consistent with the transcriptome results of butanol stress on *E. coli*[27]. A large portion (~ 30 %) of the $BtOH^-$ and $BtOH^+$ observed proteomes fall into the collective COG class of S (unknown functions), R (general function prediction only), or X (no matching COGs). Excluding these three COG classes, a large number of the observed proteins fall into COG classes J, E and G, which correspond to protein synthesis, amino acid transport and metabolism, and carbohydrate transport and metabolism, respectively. Detailed analyses of metabolic pathways using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database revealed that the enzymes involved in central metabolic pathways such as glycolysis, the tricarboxylic acid cycle and pentose phosphate pathway were all expressed under both $BtOH^-$ and $BtOH^+$ conditions.

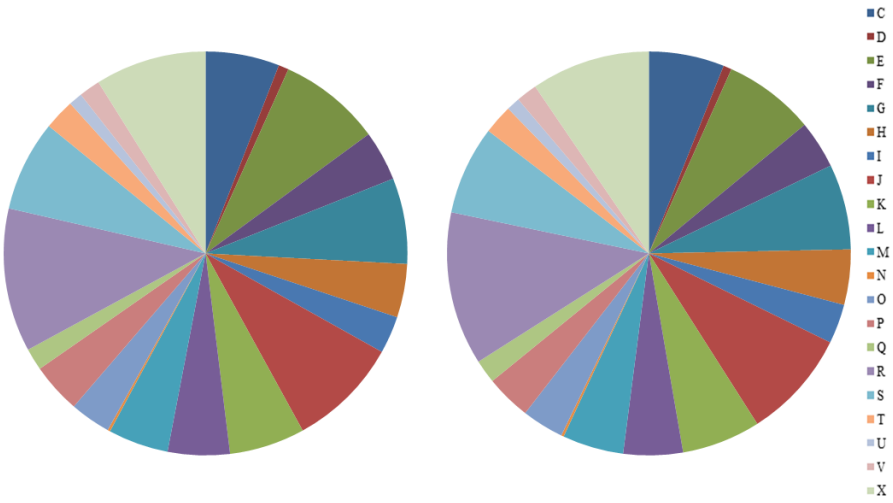


Figure 2.6 Distribution of proteins by cluster of orthologous groups showing the $BtOH^-$ (left) and $BtOH^+$ (right) proteomes of SG1. Identified proteins were from merged result of biological triplicate experiments. COGs, starting from the midnight position: C, energy production and conversion; D, cell cycle control, cell

division, chromosome partitioning; E, amino acid transport and metabolism; F, nucleotide transport and metabolism; G, carbohydrate transport and metabolism; H, coenzyme transport and metabolism; I, lipid transport and metabolism; J, translation, ribosomal structure and biogenesis; K, transcription; L, replication, recombination and repair; M, cell wall/membrane/envelope biogenesis; O, posttranslational modification, protein turnover, chaperones; P, inorganic ion transport and metabolism; Q, secondary metabolites biosynthesis, transport and catabolism; R, general function prediction only; S, function unknown; T, signal transduction mechanisms; U, intracellular trafficking, secretion, and vesicular transport; V, defense mechanisms; X, no matching COG.

2.3.3 Comparison of identified proteome and theoretical proteome

Combining all the identified proteins in $BtOH^-$ and $BtOH^+$, we were able to identify a total of 1567 unique proteins representing approximately 64 % of a total of 2457 theoretical ORFs. Although not all the predicted ORFs are necessarily present in the cell or present under our culture conditions, we still want to find out if the 36 % of un-identified ORFs have similar physical and chemical properties. Table 2.4 shows the cellular location comparison among identified proteome and the predicted ORFs. Despite a similar percentage of membrane proteins identified in SG1 grown in $BtOH^-$ and $BtOH^+$ media, our technique has detection bias towards cytoplamic proteins rather than membrane proteins. Since we used whole cell lysate without cellular fractionation, it was not surprising that membrane proteins will be under-detected given that membrane proteins are hard to solubilize and digest compared to cytoplasmic proteins. Figure 2.7 shows the molecular weight distribution of observed and theoretical proteomes. It appears that the low molecular weight proteins were under-detected compared to large proteins. One reason is that large proteins have better chance to produce more peptides and therefore have a better chance to be detected. In the future, membrane proteins and proteins in low mass region could be enriched by using cellular fractionation and low mass cut-off filters to increase proteome coverage.

Interestingly, the biggest difference between the predicted and observed proteomes is the presence of proteins that belong in COG X (Figure 2.8). This implies that many of the unidentified proteins were actually hypothetical proteins, and their expression in current growth condition is uncertain.

Table 2.4 Distribution of membrane and soluble proteins predicted and observed in the MS studies.

	# Soluble Proteins	# Membrane Proteins	% Membrane Proteins
Predicted	1842	615	25.0
Total observed	1317	250	16.0
$BtOH^-$ identified	1165	201	14.7
$BtOH^+$ identified	1180	198	14.4

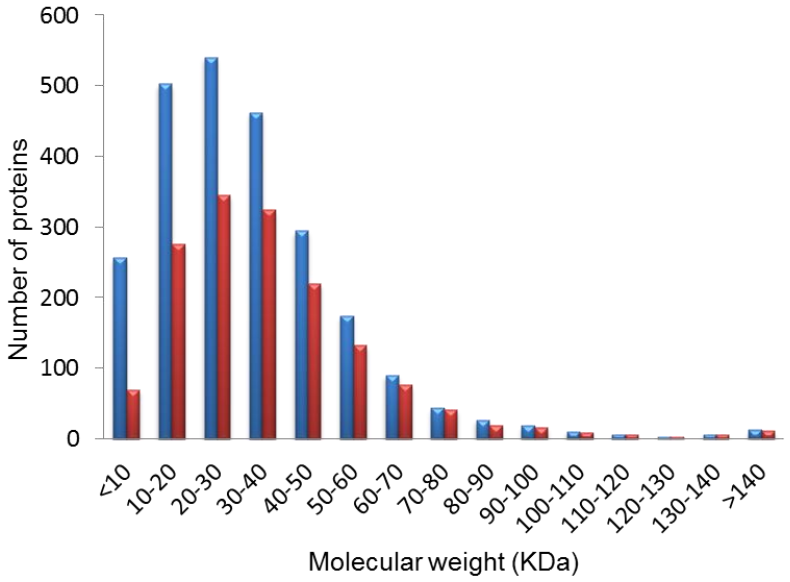


Figure 2.7 Molecular weight distribution of theoretical proteome (blue) and identified proteome (red) of *Staph. warneri* SG1. Protein identification list was generated by merging data from biological triplicates of SG1 grown in $BtOH^-$ and $BtOH^+$ media.

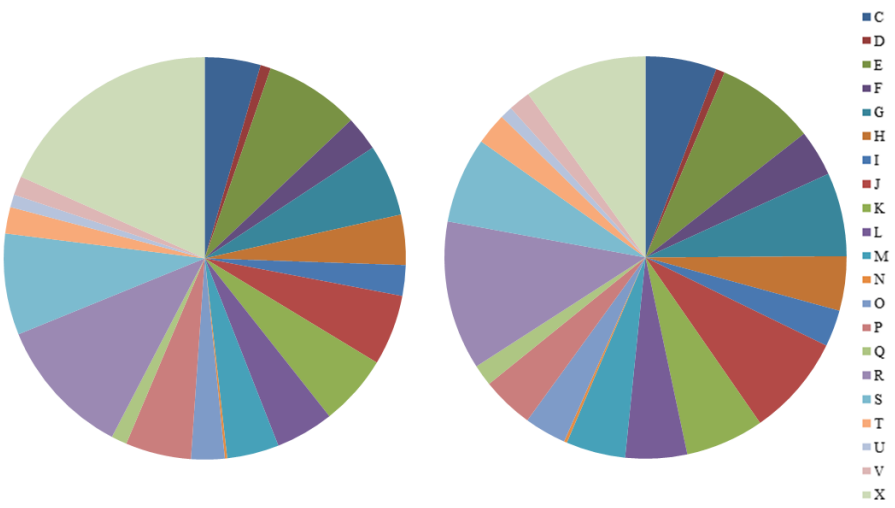


Figure 2.8 Distribution of proteins by cluster of orthologous groups showing the theoretical (left) and identified (right) proteomes of *Staph. warneri* SG1. Theoretical proteome was based on predicted ORFs and identified proteome was based on merged list of all the identified proteins from biological triplicates of $BtOH^-$ and $BtOH^+$ proteome. COG groups are as in Figure 2.6.

2.3.4 Rough quantification using exponentially modified protein abundance index

emPAI, the number of identified peptides divided by number of theoretical tryptic peptides, has proven to be a useful tool for estimating absolute protein content from the LC-MS/MS data of complex protein mixtures[19]. In this work, we directly extracted emPAI values (with a 95 % confidence threshold) from the profiling experiment using MASCOT Server. The extracted emPAI values of proteins identified in $BtOH^-$ and $BtOH^+$ were then normalized by total emPAI value of all the proteins identified in $BtOH^-$ and $BtOH^+$ respectively to avoid bias from data acquisition, mass spectrometry, and other technical variation. The normalized emPAI values were then analyzed by Metaboanalyst (http://www.metaboanalyst.ca). Commonly detected proteins were compared by fold change which was calculated by dividing the normalized emPAI value of the protein under the $BtOH^+$ condition by that from the $BtOH^-$ condition. In total, 117

observed proteins showed > 1.5-fold change and a *p*-value smaller than 0.05 (Figure 2.9)

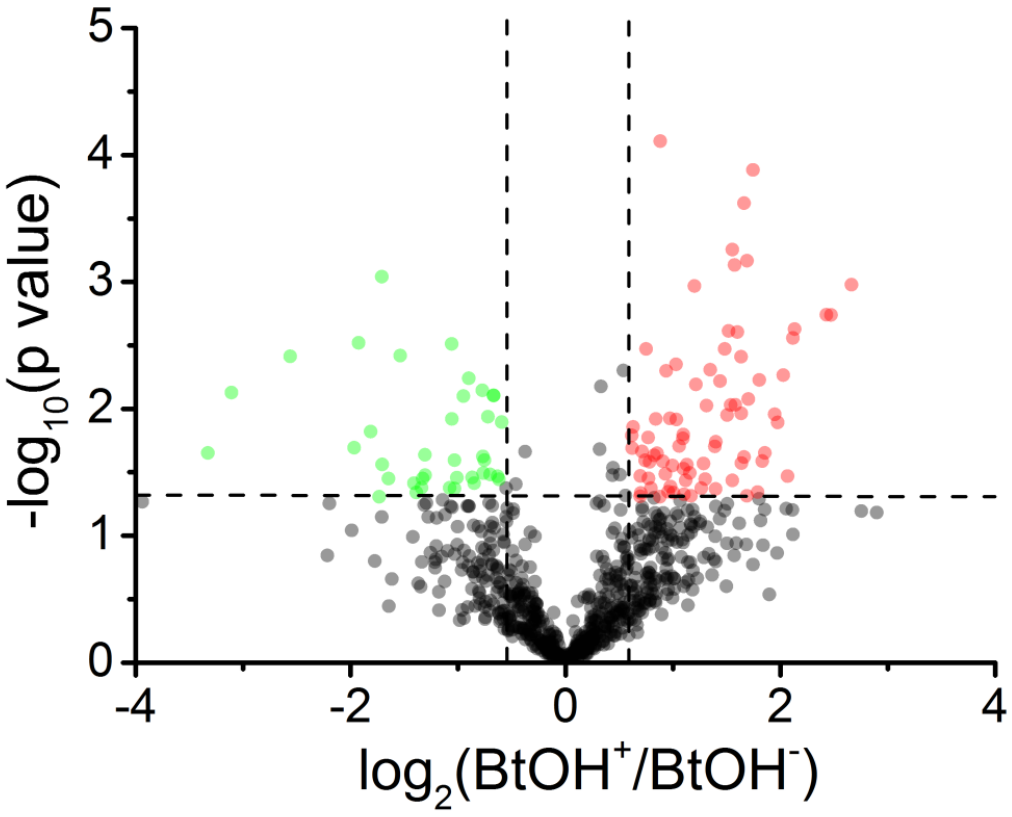


Figure 2.9 Volcano plot representing changes in protein expression levels upon butanol challenge of *Staph. warneri* SG1. Differentially expressed proteins that were up-regulated or down-regulated by at least 1.5-fold, with p-values smaller than 0.05, are marked in red and green, respectively.

2.3.5 Global expression change in SG1 upon butanol stress

To understand the physiology relevant to butanol adaption, we need to evaluate the global cellular changes of SG1 upon butanol stress. To facilitate data interpretation, we considered proteins that were only observed under either $BtOH^-$ or $BtOH^+$ condition to be extreme cases of differential expression. Thus, proteins

that were observed only in the presence of butanol are considered as highly up-regulated and vice versa. We then classified proteins in SG1 proteome by Clusters of Orthologous Groups (COGs)[26].

The global cellular changes in SG1 upon butanol challenge are represented in Figure 2.10. In the $BtOH^+$ environment, proteins in COG classes G (carbohydrate transport and metabolism), O (post-translational modification/protein turnover/chaperones), H (Coenzyme transport/metabolism), Q (Secondary metabolites biosynthesis, transport/catabolism), I (lipid transport/metabolism), M (cell wall/membrane biogenesis) and C (Energy production and conversion) tended to be up-regulated. On the other hand, proteins in COG classes P (inorganic ion transport and metabolism) and J (translation) were negatively correlated with butanol exposure.

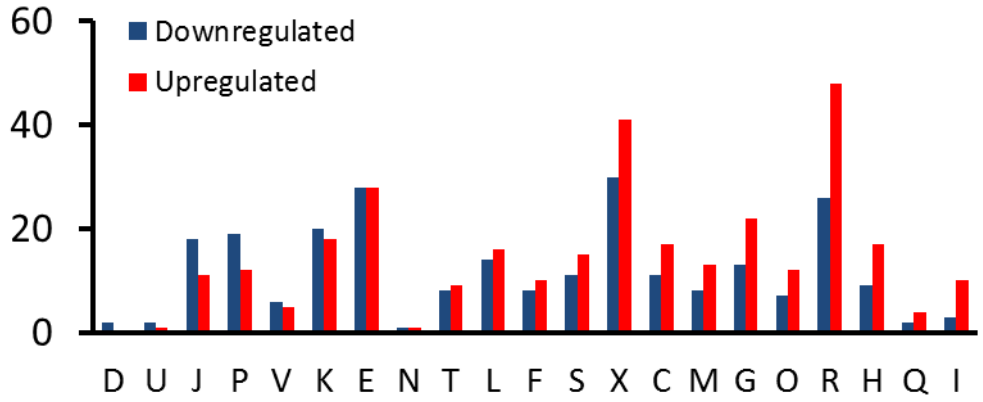


Figure 2.10 Distribution of differentially expressed proteins by cluster of orthologous groups. The COG classes were organised from the most down-regulated class to the most up-regulated class from left to right. COG groups are as in Figure 2.6.

Cell envelope

The cell envelope of bacteria is comprised of a cell membrane and a cell wall, and plays a critical role in the survival and adaptation of the cell. Research has shown that the presence of butanol can interrupt the function of bacterial cell

envelope by changing membrane physiological properties, causing misfolding of membrane proteins and by inhibiting many transport functions[1,2,4,11,23,27,28].

Interestingly, the proteins in COG M (Cell wall/membrane/envelope biogenesis) and I (Lipid transport/ metabolism) were up-regulated under butanol stress. UDP-N-acetylglucosamine 1-carboxyvinyltransferase (AGC90114), the enzyme responsible for catalyzing the first committed step of cell wall biosynthesis, was up-regulated (3.0 fold). Penicillin-binding protein 4 (AGC91338), a membrane-associated protein that catalyze the final step of murein biosynthesis, was detected only when SG1 was grown in the presence of butanol. The up-regulation of the above two enzymes suggests there was an increased rate of cell wall biosynthesis under butanol challenge. The up-regulation of penicillin-binding protein 4 might also suggest the cells were trying to alter the density of cell wall to prevent permeation of butanol[29].

Heat shock proteins and other stress response proteins

It has been previously reported that heat shock proteins play an important role in solvent tolerance of both Gram-positive and Gram-negative strain[22,28,30-32]. In this study, we identified three heat shock proteins: GroL (AGC90178, 2.0 fold), ClpC (AGC91458, 2.5 fold) and ClpX (AGC90399, 3.1 fold). They are all up-regulated in butanol stress. It has been reported that the Clp proteases are also activated upon solvent stress in the strict anaerobe *C. acetobutylicum*[11,33,34]. GroL belongs to *rpoH* regulon which has been found up-regulated in the presence of several different alcohols[27,35]. Butanol inside the cell can affect protein folding by disrupting the hydrogen bond network and hydrophobic interaction of proteins, thus the up-regulation of protein folding chaperone such as GroL might be necessary to maintain protein folding in the presence of butanol. ClpC and ClpX belong to the Clp superfamily of proteases. Interestingly, a short chain alcohol dehydrogenase (AGC91646) was expressed only in the presence of butanol. In some bacteria, short chain alcohol dehydrogenases have activities using butanol as

substrate[36]. It is not clear whether AGC91646 can catalyse oxidation of butanol, but the up-regulation of this protein could be part of detoxification mechanism for SG1 to metabolically break down butanol molecules inside the cells.

Energy metabolism

It has been reported that a high concentration of butanol inhibits active nutrient transport, the membrane bound ATPase and glucose uptake[2,28]. Thus not surpringly, we would expect energy-dependent processes, as well as those that generate energy (i.e. ATP generation) would be largely affected in SG1 during butanol stress. In agreement, proteins involved in energy production and conversion (COG C) as well as carbohydrate transport and metabolism (COG G) were significantly up-regulated when the cells were exposed to butanol. Among them, enzymes in glycolytic pathway such as pyruvate kinase (AGC90378), glucose-6-phosphate dehydrogenase (AGC90555), L-lactate dehydrogenase (AGC89629), aldehyde dehydrogenase (AGC90092) and dihydrolipoyl dehydrogenase (AGC90987) were all up-regulated. No proteins in the glycolytic pathway are found to be down-regulated. We also observed an increase in expression of glucose-specific IIB component in phosphotransferase system (AGC89683, 1.8 fold) which belongs to a complex group translocation system for sugar uptake. These observations suggest that processes which are involved in glucose uptake and energy generation are up-regulated to compensate those energy-consuming processes for combating butanol stress, which is in line with studies in *C. acetobutylicum*[37,38]. Interestingly, we observed nine ribosomal proteins which were down-regulated in butanol stress. This result is similar to the study on solvent stress of the cyanobacteria *Synechocystis* PCC 6803[39]. Down-regulation of ribosomal proteins suggests decreased activities of protein biosynthesis and a potential slowdown of metabolism which might in turn cut down energy consumption.

2.4 Conclusion

The 2D-LC-MS/MS shotgun approach is a very powerful method for observing global cellular events by measuring a large set of gene expression products. In this study, we employed this particular approach to establish proteome reference maps of SG1 with and without butanol stress. The proteome reference maps allowed us to examine protein expression on a semi-quantitative basis using emPAI and to better understand the butanol tolerance phenotype of SG1. Comparative proteomics analysis of SG1 grown in the presence and absence of 1.5 % butanol revealed 117 differentially expressed proteins, 201 unique proteins under the $BtOH^+$ condition, and 189 proteins uniquely expressed under $BtOH^-$ condition. The up-regulated proteins are mainly involved in protein folding, energy metabolism, cell envelope biosynthesis as well as oxidative stress response. Meanwhile, a great portion of the down-regulated proteins observed are involved in translation and protein synthesis. The results suggest that SG1 might have developed a comprehensive mechanism for butanol-tolerance.

Supporting information available

The list of identified proteins of *Staph. warneri* SG1 cultured in $BtOH^-$ and $BtOH^+$ media and the validated MS/MS spectra library of SG1 are stored in a hard disk in Dr. Liang Li's lab and is available upon request.

2.5 Literature cited

(1) Baer, S. H.; Blaschek, H. P.; Smith, T. L. *Applied and Environmental Microbiology* **1987**, *53*, 2854.

(2) Bowles, L. K.; Ellefson, W. L. *Applied and Environmental Microbiology* **1985**, *50*, 1165.

(3) Casal, M.; Cardoso, H.; Leao, C. *Applied and Environmental Microbiology* **1998**, *64*, 665.

(4) Vollherbst-Schneck, K.; Sands, J. A.; Montenecourt, B. S. *Applied and Environmental Microbiology* **1984**, *47*, 193.

(5) Sardessai, Y.; Bhosle, S. *Research in Microbiology* **2002**, *153*, 263.

(6) Cheng, V. W.; Zhang, G.; Oyedotun, K. S.; Ridgway, D.; Ellison, M. J.; Weiner, J. H. *Genome Announcements* **2013**, *1*, e0003813.

(7) Pandey, A.; Mann, M. *Nature* **2000**, *405*, 837.

(8) Gygi, S. P.; Rochon, Y.; Franza, B. R.; Aebersold, R. *Molecular and Cellular Biology* **1999**, *19*, 1720.

(9) Hochstrasser, D. F.; Sanchez, J. C.; Appel, R. D. *Proteomics* **2002**, *2*, 807.

(10) Mao, S.; Luo, Y.; Bao, G.; Zhang, Y.; Li, Y.; Ma, Y. *Molecular BioSystems* **2011**, *7*, 1660.

(11) Mao, S.; Luo, Y.; Zhang, T.; Li, J.; Bao, G.; Zhu, Y.; Chen, Z.; Zhang, Y.; Li, Y.; Ma, Y. *Journal of Proteome Research* **2010**, *9*, 3046.

(12) Wolff, S.; Otto, A.; Albrecht, D.; Zeng, J. S.; Büttner, K.; Glückmann, M.; Hecker, M.; Becher, D. *Molecular & Cellular Proteomics* **2006**, *5*, 1183.

(13) Otto, A.; Bernhardt, J.; Meyer, H.; Schaffer, M.; Herbst, F.-A.; Siebourg, J.; Mäder, U.; Lalk, M.; Hecker, M.; Becher, D. *Nature Communications* **2010**, *1*, 137.

(14) Unell, M.; Abraham, P. E.; Shah, M.; Zhang, B.; Rückert, C.; VerBerkmoes, N. C.; Jansson, J. K.; Ru, C. *Journal of Proteome Research* **2009**, *8*, 1953.

(15) Ye, X.; Li, L. *Analytical Chemistry* **2012**, *84*, 6181.

(16) Wang, N.; Xie, C.; Young, J. B.; Li, L. *Analytical Chemistry* **2009**, *81*, 1049.

(17) Wang, N.; Li, L. *Analytical Chemistry* **2008**, *80*, 4696.

(18) Elias, J. E.; Gygi, S. P. *Nature Methods* **2007**, *4*, 207.

(19) Ishihama, Y.; Oda, Y.; Tabata, T.; Sato, T.; Nagasu, T.; Rappsilber, J.; Mann, M. *Molecular & Cellular Proteomics* **2005**, *4*, 1265.

(20) Xu, M.; Li, L. *Journal of Proteome Research* **2011**, *10*, 3632.

(21) Segura, A.; Godoy, P.; van Dillewijn, P.; Hurtado, A.; Arroyo, N.; Santacruz, S.; Ramos, J. L. *Journal of Bacteriology* **2005**, *187*, 5937.

(22) Wijte, D.; van Baar, B. L.; Heck, A. J.; Altelaar, A. F. *Journal of Proteome Research* **2011**, *10*, 394.

(23) Sivagnanam, K.; Raghavan, V.; Shah, M.; Hettich, R.; Verberkmoes, N.; Lefsrud, M. *Proteome Science* **2011**, *9*, 66.

(24) Schwartz, R.; Ting, C. S.; King, J. *Genome Research* **2001**, *11*, 703.

(25) Mastroleo, F.; Leroy, B.; Van Houdt, R.; s' Heeren, C.; Mergeay, M.; Hendrickx, L.; Wattiez, R. *Journal of Proteome Research* **2009**, *8*, 2530.

(26) Tatusov, R. L.; Galperin, M. Y.; Natale, D. A.; Koonin, E. V. *Nucleic Acids Research* **2000**, *28*, 33.

(27) Rutherford, B. J.; Dahl, R. H.; Price, R. E.; Szmidt, H. L.; Benke, P. I.; Mukhopadhyay, A.; Keasling, J. D. *Applied and Environmental Microbiology* **2010**, *76*, 1935.

(28)Nicolaou, S. a.; Gaida, S. M.; Papoutsakis, E. T. *Metabolic Engineering* **2010**, *12*, 307.

(29) Huffer, S.; Clark, M. E.; Ning, J. C.; Blanch, H. W.; Clark, D. S. *Applied and Environmental Microbiology* **2011**, *77*, 6400.

(30) Mastroleo, F.; Leroy, B.; Houdt, R. V.; Heeren, C.; Mergeay, M.; Hendrickx, L.; Wattiez, R. *Journal of Proteome Research* **2009**, 2530.

(31) Nicolaou, S. A.; Gaida, S. M.; Papoutsakis, E. T. *Metabolic Engineering* **2010**, *12*, 307.

(32) Segura, A.; Molina, L.; Fillet, S.; Krell, T.; Bernal, P.; Munoz-Rojas, J.; Ramos, J. L. *Current Opinion in Biotechnology* **2012**, *23*, 415.

(33) Tomas, C. A.; Beamish, J.; Papoutsakis, E. T. *Journal of Bacteriology* **2004**, *186*, 2006.

(34) Alsaker, K. V.; Paredes, C.; Papoutsakis, E. T. *Biotechnology and Bioengineering* **2010**, *105*, 1131.

(35) Brynildsen, M. P.; Liao, J. C. *Molecular Systems Biology* **2009**, *5*, 277.

(36) Walter, K. A.; Bennett, G. N.; Papoutsakis, E. T. *Journal of Bacteriology* **1992**, *174*, 7149.

(37) Alsaker, K. V.; Paredes, C.; Papoutsakis, E. T. *Biotechnology and Bioengineering* **2010**, *105*, 1131.
(38) Alsaker, K. V.; Spitzer, T. R.; Eleftherios, T.; Papoutsakis, E. T. *Journal of Bacteriology* **2004**.
(39) Liu, J.; Chen, L.; Wang, J.; Qiao, J.; Zhang, W. *Biotechnology for Biofuels* **2012**, *5*, 68.

Chapter 3. Quantitative Proteomic and Metabolomic Analysis of *Staphylococcus warneri* SG1 Cultured in the Presence and Absence of Butanol

3.1 Introduction

Engineering solvent tolerant microorganisms for biodegradation, biofuel production, and biocatalysis of high value compounds is an important branch of synthetic biology. However, these endeavors are often hindered by the toxicity of organic compounds, which damage important macromolecules such as DNA, RNA, and proteins, as well as disrupt biological membrane functions such as transport and dissipation of the proton motive force. Adaptation of the bacterium to survive in a high titer of organic solvent is achieved through global changes that include alteration of the membrane structure and fluidity, differential protein expression, and activation of specific defense mechanisms. The interplay between solvent stress and cellular response is thoroughly reviewed[1-6]. In spite of this, the fact that solvent tolerant bacteria isolated, either selectively or naturally, outnumbers those with acquired tolerance via targeted genetic manipulation is a testament of the complexity and multi-facet nature of how microorganisms cope with solvent stress.

In general, Gram-negative bacteria have a higher tolerance against organic solvents because they have an additional outer membrane and thus are more widely studied compared to Gram-positive bacteria[3]. In recent years, the number of solvent tolerant Gram-positive bacteria isolated and studied has increased, especially those from the *Staphylococcus* and *Bacillus* genera[7-10]. *Staphylococcus warneri* is a solvent tolerant Gram-positive bacterium that constitutes a part of the human skin flora, and its genome was recently published[11]. As described in Chapter 2, proteomic profiling using the 2D-LC-MS/MS shotgun approach

*Victor Cheng contributed to the cell culture, data processing and interpretation of this work (biological characterization part). Yiman Wu contributed partially to the sample preparation and data analysis of the metabolomics work.

covered 64 % of theoretical proteome of SG1, and rough quantification analysis indicates a complicated mechanism explaining butanol tolerance of SG1.

Classical fermentation of *Clostridium acetobutylicum* on molasses or grains yields acetone, butanol, and ethanol[12]. With the advent of synthetic biology, notably in *Escherichia coli* and *Saccharomyces cerevisiae*, novel or heterologous metabolic pathways can be engineered to produce butanol and numerous high value chemicals[13-16]. However, butanol titers from these biological systems are believed to be limited by the chaotropic effects of the end product. A number of studies on the aforementioned model organisms have attempted to elucidate the complex mechanism of solvent tolerance using genomics, proteomics, and metabolomics discovery approaches[17-21]. Though informative, the shotgun proteomic approach identified only a small subset of predicted proteins using *C. acetobutylicum*[20,21] and *Pseudomonas putida*[22,23] as query organisms.

In this study, we employed 2-MEGA labeling to carry out quantitative proteomic analysis of SG1 grown in the presence and absence of 1-Butanol using 2D-LC-MS/MS. Of the 967 quantified proteins, we found proteins involved in energy metabolism, lipid and cell envelope biogenesis, and those with chaperone functions to be differentially up-regulated. Finally we used an isotope labeling LC-MS method to investigate the metabolomic changes of SG1 upon 1-Butanol exposure. The combination of proteomic and metabolomic data provides detailed insight into the solvent tolerant mechanism of SG1.

3.2 Experimental Procedures

3.2.1 Chemicals and reagents.

All the chemicals and reagents, unless otherwise stated, were purchased from Sigma-Aldrich Canada (Markham, ON, Canada). Lysostaphin was purchased from AMBI Products (Lawrence, NY). Isotopically enriched reagents were purchased from Cambridge Isotope Laboratories (Andover, MA). Phosphoric acid

(H_3PO_4), potassium chloride (KCl), potassium dihydrogen phosphate (KH_2PO_4), and ammonium bicarbonate (NH_4HCO_3) were purchased from EMD Chemical, Inc. (Mississauga, ON, Canada). Sequencing grade modified trypsin, HPLC grade formic acid, LC-MS grade water, acetone, and acetonitrile (ACN) were purchased from Fisher Scientific Canada (Edmonton, AB, Canada). A domestic 900 W (2450 MHz) sunbeam microwave oven was used to perform microwave-assisted protein solubilization experiments.

3.2.2 Cell growth and protein sample preparation.

Cultures of *Staphylococcus warneri* strain SG1 were routinely grown in Luria Bertani broth at 37 ^{o}C with shaking for 16 hours. For MS studies, 2 L cultures were grown in triplicate (seeded with a 0.1 % inoculum), with or without 1.5 % (V/V) 1-Butanol, and harvested by centrifugation at 8,000 x *g* for 15 minutes and resuspended in 100 mM Tris / 5 mM EDTA buffer, pH 7.0. Cell lysis was carried out either mechanically by repeated passage (4x) through a Constant Systems TS benchtop cell disruptor (Daventry, Northants., UK) at 40 kpsi, or enzymatically by adding NaCl (100 mM), lysostaphin (10 μgmL^{-1}) and lysozyme (50 μgmL^{-1}), followed by incubation at 37 ^{o}C for 1 hour. Unbroken cells and cell debris were removed by centrifugation at 10,000 x *g* for 20 minutes, and the cell lysates were frozen immediately with liquid nitrogen.

Protein assays were performed to adjust protein concentrations of lysates to similar levels. Acetone, pre-cooled to -80 ^{o}C, was gradually added to the whole cell lysates to a final concentration of 80 % (V/V) and the mixtures were incubated overnight at -80 ^{o}C. Samples were then spun at 20,800 × *g* for 20 minutes and the pellets were washed with 40 μL of pre-chilled acetone before drying at room temperature. The pellets were then subjected to microwave-assisted protein solublization in urea[24]. Briefly, 8 M urea was added to the whole cell lysates and microwave irradiation was applied for 6 times in 30 s cycles with sample cooling and homogenization between cycles. The mixtures were then

centrifuged at 20,800 × *g* for 5 minutes and the pellets were subjected to a fresh round of microwave assisted urea extraction. Upon complete solubilization of the pellets, the supernatant fractions were pooled and diluted with 100 mM NH_4HCO_3 to reduce the urea concentration to ~1 M. Samples were analyzed by protein assay and reduced with dithiothreitol for 1 h at 37 ^oC, followed by alkylation with iodoacetamide for 0.5 h at room temperature in the dark. Trypsin was added to a protein:tryspin ratio of 40:1 and incubated at 37 ^oC for 20 h for complete digestion. The tryptic digests were acidified with 50 % trifluoroacetic acid to pH 2 and injected into Agilent 1100 HPLC system (Palo Alto, CA) for desalting and quantification. A Polaris C18-A column (4.6 mm × 50 mm, 3 μm particle size, 300 Å pore size) (Varian, CA) was used for desalting and a UV detector operating at 214 nm was used for quantification of the eluted peptides[25].

3.2.3 Quantification using 2-MEGA labeling and 2D-LC-MS/MS.

2-MEGA isotopic labeling was carried out on biological triplicate of SG1 grown in the absence and presence of BtOH according to the workflow shown in Figure 3.1. After protein digestion, the $BtOH^-$ and $BtOH^+$ peptides were individually labeled with either heavy chain or light chain using the 2-MEGA labeling method[26]. Briefly, 4 M O-methylisourea was added to the peptide mixtures and the pH was adjusted to 11 with 2 M NaOH. Samples were incubated at 60 °C for 20 min with intermittent shaking to guanidinate the lysine side chains. The pH was then adjusted to 6 using 50 % trifluoroacetic acid and further adjusted to 4.5 using acetate buffer. 1 M $NaCNBH_3$ and 4 % formaldehyde ($^{12}CH_2O$ for light chain labeling and $^{13}CD_2O$ for heavy chain labeling) were added to dimethylate the N-termini of the peptides. After labeling, a small amount of 1 M NH_4HCO_3 was added to consume the excess formaldehyde and the reaction was quenched by adjusting the pH to 2 using 10 % trifluoroacetic acid. Finally, labeled peptides were desalted and quantified. Heavy chain labeled $BtOH^-_H$ was mixed with light chain labelled $BtOH^+_L$ in a 1:1 ratio based on the total peptide content by weight as forward labeling. Similarly, reverse labeling was done by

mixing light chain labeled $BtOH^-_L$ with heavy chain labeled $BtOH^+_H$. The $BtOH^-_L$:$BtOH^+_H$ and $BtOH^+_L$:$BtOH^-_H$ mixtures were analyzed by 2D-LC-MS/MS.

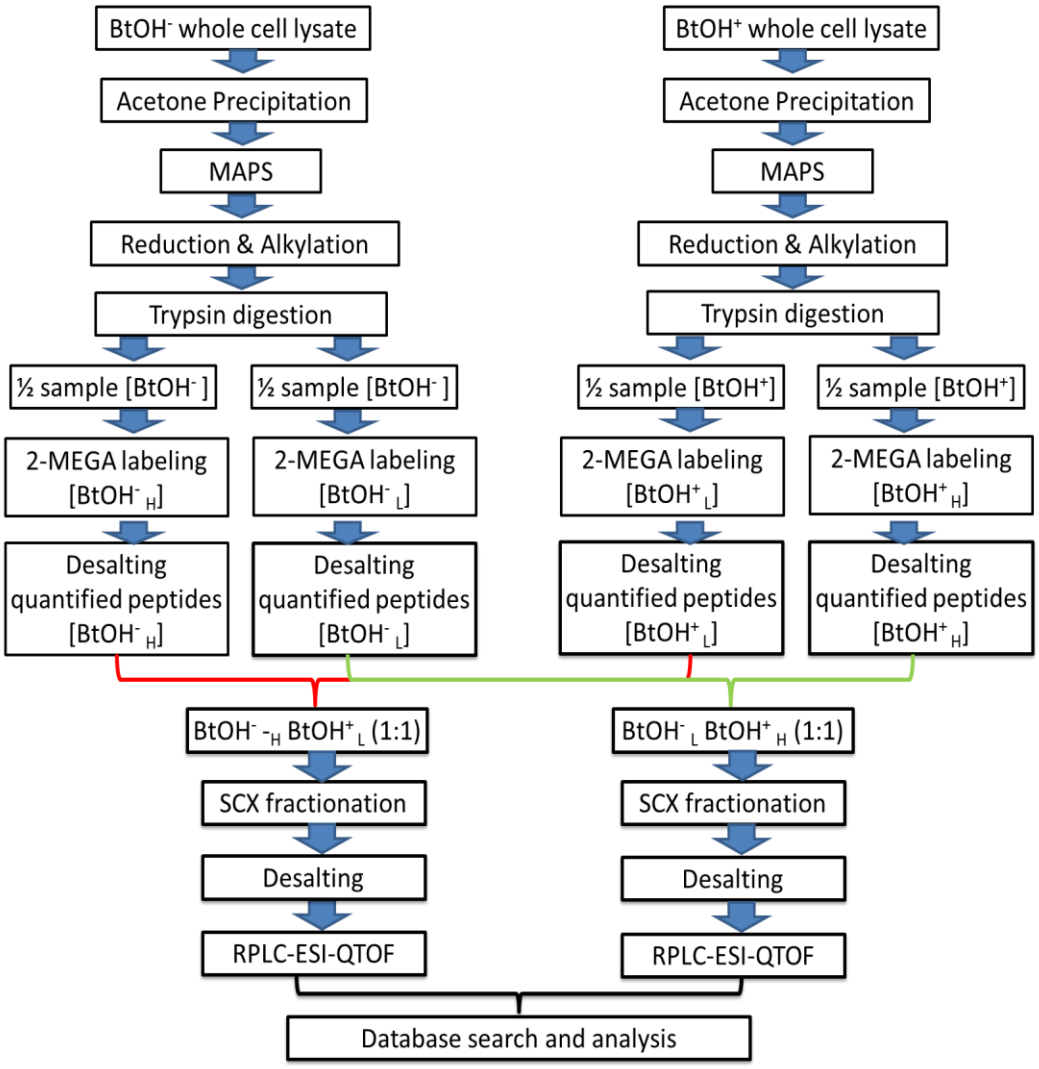


Figure 3.1 Experimental workflow of 2-MEGA isotopic labeling experiment for quantitating changes of protein expression in the $BtOH^-$ and $BtOH^+$ proteomes.

The mixed peptides were first separated by SCX liquid chromatography using a polysulfoethyl A column (2.1 mm x 250 mm, 5 μm particle size, 300 Å pore size) (PolyLC, Columbia, MD). Peptides were separated into 20 fractions using the following gradient: mobile phases A (10 mM KH_2PO_4, pH 2.76) and B

(10 mM KH_2PO_4, pH 2.76, 500 mM KCl), t = 0 min, 0 % B; t = 1 min, 4 % B; t = 17 min, 20 % B; t = 39 min, 60 % B; t = 45 min, 100 % B; t = 50 min, 100 % B; t = 52 min, 0 % B; t = 62 min, 0 % B. The collected peptide fractions were then desalted and quantified. Less abundant neighbor fractions were combined to a total of 20 fractions.

The SCX fractionated peptides were further separated by reversed phase liquid chromatography (RPLC) using a nanoACQUITY UltraPerformance LC system (Waters, Missisauga, ON) with an Atlantis dC18 column (75 μm × 150 mm, 3μm particle size, 100 Å pore size) (Waters, Milford, MA). The following gradient was applied to separate the peptides: mobile phases A (0.1 % FA in water) and B (0.1 % FA in ACN); t = 0 min, 2 % B; t = 2 min, 7 % B; t = 85 min, 20 % B; t = 105 min, 30 % B; t = 110 min, 45 % B; t = 120 min, 90 % B; t = 125 min, 90 % B; t = 130 min, 95 % B[25]. The eluted peptides were then electrosprayed into an electrospray ionization (ESI) quadrupole time-of-flight (Q-TOF) premier mass spectrometer (Waters, Missisauga, ON) at a flow rate of 350 $nLmin^{-1}$. A survey MS scan was acquired from m/z 350-1600 for 0.8 s, followed by 4 data-dependent MS/MS scans. A mass tolerance window of 80 mDa was applied for both dynamic and precursor ion exclusion[27], with a retention time tolerance window of 150 s. The collision energy used for MS/MS analysis was varied based on the precursor ion mass and charge state. A mixture of leucine enkephalin and (Glu1)-fibrinopeptide B, used as mass calibrants (i.e., lock-mass), was infused at a flow rate of 300 $nLmin^{-1}$, and a 1 s MS scan was acquired every 1 min throughout the run. Each SCX fraction was analyzed once on the RPLC-MS with a precursor ion exclusion list generated from neighbour fractions to eliminate redundant identification.

Raw MS and MS/MS data were searched using MASCOT DISTILLER with the following parameters: taxonomy, all entries; enzyme, trypsin; missed cleavages, 1; fixed modifications, carbamidomethylation (C) and guanidination (K); variable modifications, dimethylation d_0 ($+^{12}C_2H_4$, +28.0313 Da, N-term),

dimethylation d_6 ($+^{13}C_2D_4$, +34.0631 Da, N-term); Oxidation (M); peptide tolerance, 0.2 Da; MS/MS tolerance, 30 ppm. A modified ESI-Q-TOF ion fragmentation series that permitted a-type ions was applied. The relative intensity ratios for peak pairs were extracted and normalized using the median ratios to avoid system bias. In cases where same peptide pairs were detected multiple times in different SCX fractions or where the peptide pairs with the same sequences but different charge states, the relative ratios of the peptides pairs were averaged. Peptides with relative error >1 in forward and reverse labeling were discarded. Protein ratios were calculated based on the geometric mean of peptide ratios from the same protein. Finally, six lists of protein ratios were generated from the biological triplicate samples (each sample set contains a pair of forward and reverse labeling data). To identify differentially expressed proteins with statistic confidence, the following procedures were used. First, proteins quantified in less than two biological samples were discarded. The geometric mean of protein ratios from different replicates were calculated and subjected to one-sample *t*-test where only proteins with $p<0.01$ were retained[28,29], followed by applying a 1.5 fold cut-off threshold for differential expression. Global protein expression profiles were analyzed by Cluster of orthologous groups (COG)[30] and the Kyoto Encyclopedia of Genes and Genomes (KEGG) server[31].

3.2.4 Metabolomic analysis on amine- containing metabolites and TCA cycle metabolites

For metabolite extraction, whole cell lysates from $BtOH^-$ and $BtOH^+$ cells were prepared as described above. 1200 μL of acetone was added to 300 μL of the whole cell lysates to precipitate the proteins. The supernatants were dried using a SpeedVac and resuspended in either 50 μL of water for dansylation labeling or 90 μL of water for *p*-dimethylaminophenacyl (DmPA) labeling.

For dansylation labeling, the 50 μL solution was mixed with 25 μl 250 mM sodium carbonate/sodium bicarbonate buffer and ACN, vortexed, spun down and

mixed with 50 μL freshly prepared ^{12}C-dansyl chloride solution (18 mg mL^{-1}) (light labeling) or ^{13}C-dansyl chloride solution (18 mg mL^{-1}) (heavy labeling). The reaction was allowed to proceed for 1 hour at 60 °C, followed by addition of 10 μL 250mM NaOH to quench the excess dansyl chloride. The solution was then incubated at 60°C for another 10 min. Finally, 50μL 425mM formic acid in 1:1 ACN:H_2O was added to consume the excess NaOH and to acidify the solution

For DmPA labeling, 90 μL solution was first acidified with HCl and extracted with 300 μL of ethyl acetate. The organic layer was dried and dissolved in 60 μL 20 mg mL^{-1} triethylamine, then mixed with 60 μL freshly prepared ^{12}C-DmPA bromide solution (20 mg mL^{-1}) (light labeling) or ^{13}C-DmPA bromide solution (20 mg mL^{-1}) (heavy labeling). The reaction was allowed to proceed for 50 min at 90 ^{o}C and quenched with 50 μL 20 mg mL^{-1} triphenylacetic acid for 30 min at 90 ^{o}C.

An LC-UV quantification step was carried out prior to mass analysis in order to control the amount of sample used for metabolome comparison. 2 μL of the labeled solution was injected onto a Phenomenex C18 column (2.1 mm × 5 cm, 1.7 μm particle size, 100 Å pore size). For amine quantification, the gradient started with 100 % A (0.1 % (v/v) formic acid in 5 % (v/v) acetonitrile) and 0 % B (0.1 % (v/v) formic acid in acetonitrile) for 1 min and was changed to 5 % A / 95 % B within 0.01 min and held for 1 min. The gradient was restored to 100 % A / 0 % B in 0.5 min and hold at this condition for 3.5 min to re-equilibrate the column. For acid quantitation, the gradient started with 20% B for 2 min and was increased to 85% B within 0.01 min and held for 2 min and then increased to 95% B within 0.01 min and held at 95% B for 1 min. The gradient was restored back to 20% B in 1 min and held for 3 min to re-equilibrate the column. The flow rate was 450 μL min^{-1}.

The labeled metabolites were analyzed using a Bruker 9.4 Tesla Apex-Qe Fourier transform ion-cyclotron resonance (FT-ICR) mass spectrometer (Bruker,

Billerica, MA) linked to an Agilent 1100 series binary HPLC system (Agilent, Palo Alto, CA). The samples were injected onto an Agilent reversed phase Eclipse Plus C18 column (2.1 mm × 10 cm, 1.8 μm particle size, 95 Å pore size) for separation. Solvent A was 0.1 % (v/v) formic acid in 5 % (v/v) ACN, and solvent B was 0.1 % (v/v) formic acid in ACN. The chromatographic conditions for amine labeling were: t = 0 min, 20 % B; t = 3.5 min, 35 % B; t = 18 min, 65 % B; t = 21 min, 95 % B; t = 26 min, 95 % B. The gradient for acid labeling was t = 0 min, 20 % B; t = 9 min, 50 % B; t = 22 min, 65 % B; t = 26 min, 80 % B; t = 29 min, 98 % B; t = 30 min, 98 % B, The flow rate was 180 μL min^{-1}. All MS spectra were obtained in the positive ion mode. The resulting MS data were processed using R language program based on XCMS [32] written specifically for $^{12}C/^{13}C$ peak pair picking[33]. This program eliminated the false positive peaks, such as isotopic peaks, common adduct ions, and multiple charged ions. Only the protonated ion pairs were exported for further analysis.

3.3 Results and Discussion

Staphylococcus warneri strain SG1 is a solvent tolerant Gram-positive bacterium that can thrive in the presence of short-chain alcohols, alkanes, esters and cyclic aromatic compounds. The complete genome sequence of *Staphylococcus warneri* strain SG1, was recently published[11] and comparative proteomic profiling has been described in Chapter 2. Proteome profiling with high confidence enable us to establish a proteome reference map of SG1 and have an idea of the metabolic pathway of this bacteria. However, obtaining accurate quantitative information about proteome changes upon butanol challenge is more critical if we want to investigate the molecular mechanisms activated or repressed upon butanol challenge. In this study, we adopted 2-MEGA isotopic labeling in combination with 2D-LC-MS/MS approach to study the proteome expression change of SG1 grown in the absence ($BtOH^-$) and presence ($BtOH^+$) of 1.5 % 1-Butanol, a concentration which decreased cell yield at stationary phase by approximately 15 %.

3.3.1 Quantification by 2D-LC-MS/MS

We employed 2-MEGA forward and reverse labeling, followed by 2D-LC-MS/MS on SG1 grown in the presence and absence of 1-Butanol in biological triplicate. A final list of peptide pairs from both forward and reverse labeling experiments was generated. By discarding outlier data with relative error larger than 1, correlation near ideal (slope = 1 and intercept = 0) was observed in a $\log_2$-$\log_2$ plot (Figure 3.2), suggesting good analytical reproducibility between forward and reverse labeling experiments. Data from six independent labeling experiments were then integrated to give a total of 967 quantified proteins using 2-MEGA labeling. After employing a double-filter, 156 and 104 proteins were identified to be up-regulated or down-regulated at least 1.5-fold respectively, and with *p*-values smaller than 0.01 (Figure 3.3; Table 3.1). The global cellular changes in SG1 upon butanol challenge are represented by Figure 3.4. In $BtOH^+$ environment, proteins in COG classes G (carbohydrate transport and metabolism), O (post-translational modification/protein turnover/chaperones), L (replication, recombination/repair), E (amino acid transport and metabolism), I (lipid transport and metabolism) and M (cell wall / membrane biogenesis) tended to be up-regulated and accounted for more than 50 % of the 156 up-regulated proteins. On the other hand, proteins in COG classes P (inorganic ion transport and metabolism), K (transcription) and J (translation) were negatively correlated with butanol exposure.

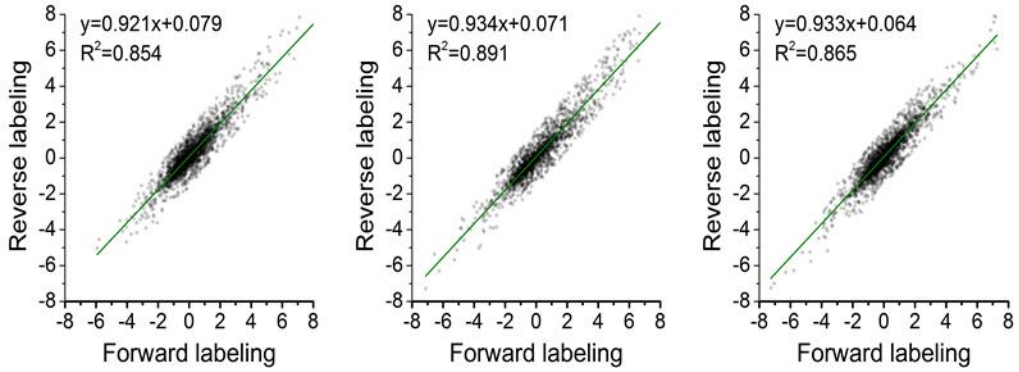


Figure 3.2 Log-log plots of peptide ratios from the forward ($BtOH^{+}_{L}$:$BtOH^{-}_{H}$) and reverse ($BtOH^{+}_{H}$:$BtOH^{-}_{L}$) 2-MEGA labeled peptides. The three graphs represent independent analyses carried out on biological triplicates of $BtOH^{-}$ and $BtOH^{+}$ samples.

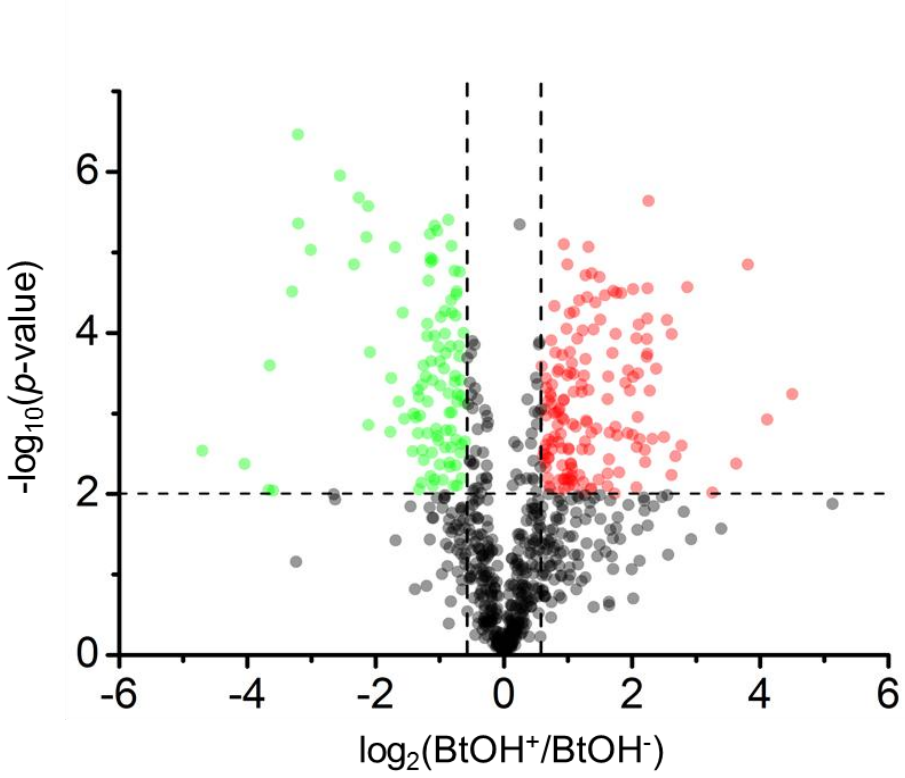


Figure 3.3 Volcano plot representing changes in protein expression levels upon butanol challenge of *Staph. warneri* SG1. Differentially expressed proteins that were up-regulated or down-regulated by at least 1.5-fold, with *p*-values smaller than 0.01, are marked in red and green, respectively.

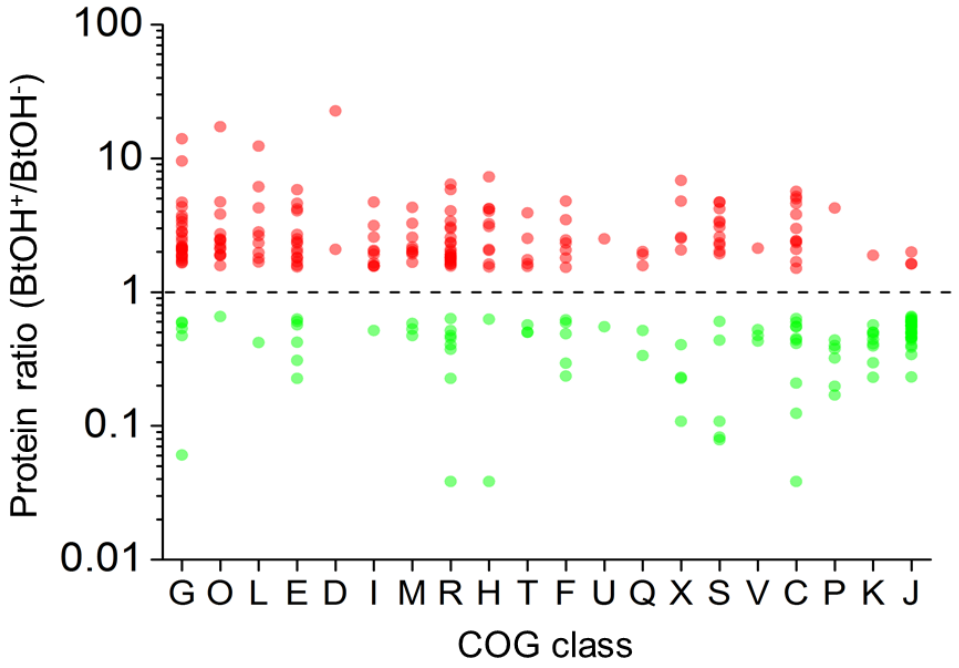


Figure 3.4 Distribution of differentially expressed proteins by cluster of orthologous groups. The COG classes were organized from the most up-regulated class to the most down-regulated class from left to right. C, energy production and conversion; D, cell cycle control, cell division, chromosome partitioning; E, amino acid transport and metabolism; F, nucleotide transport and metabolism; G, carbohydrate transport and metabolism; H, coenzyme transport and metabolism; I, lipid transport and metabolism; J, translation, ribosomal structure and biogenesis; K, transcription; L, replication, recombination and repair; M, cell wall/membrane/envelope biogenesis; O, posttranslational modification, protein turnover, chaperones; P, inorganic ion transport and metabolism; Q, secondary metabolites biosynthesis, transport and catabolism; R, general function prediction only; S, function unknown; T, signal transduction mechanisms; U, intracellular trafficking, secretion, and vesicular transport; V, defense mechanisms; X, no matching COG.

Table 3.1 List of differentially expressed proteins from *Staph. warneri* SG1 grown in $BtOH^+$ or $BtOH^-$ medium. The fold change represents the levels of protein expression change of SG1 grown in $BtOH^+$ versus $BtOH^-$. COG groups are as in Figure 3.4.

Up-regulated proteins

Fold change	Accession number	Protein name (from gene bank)	COG class	KEGG	*p*-value
5.66	AGC91120	NADH dehydrogenase-like protein	C	K03885	1.96E-03
5.20	AGC90793	succinyl-CoA synthetase subunit alpha	C	K01902	2.78E-04
4.98	AGC90962	Glycerophosphoryl diester phosphodiesterase	C	K01126	2.07E-03
4.60	AGC90092	aldehyde dehydrogenase	C	K00128	2.84E-03
3.81	AGC89629	L-lactate dehydrogenase	C	K00016	1.75E-03
2.42	AGC90442	putative oxygenase	C	N/A	1.93E-05
2.41	AGC89706	NAD(P)H-flavin oxidoreductase	C	K00540	2.23E-03
2.35	AGC90794	succinyl-CoA synthetase subunit beta	C	K01903	9.32E-05
1.69	AGC91382	aldo/keto reductase	C	N/A	4.39E-04
1.51	AGC90751	2-oxoglutarate ferredoxin oxidoreductase subunit beta	C	K00175	2.59E-04
2.09	AGC89928	glycerate dehydrogenase	CHR	N/A	4.77E-03
2.99	AGC90034	alcohol dehydrogenase	CR	K00001	3.41E-05
22.62	AGC90054	hypothetical protein	D	K03593	5.77E-04
2.09	AGC90848	cell-division initiation protein	D	K04074	6.90E-03
4.61	AGC90368	proline dipeptidase	E	K01271	4.03E-03
4.05	AGC90661	oligoendopeptidase F	E	K01417	3.51E-04
2.69	AGC90585	chorismate synthase	E	K01736	4.18E-05
2.50	AGC90525	glycine dehydrogenase subunit 2	E	K00283	8.58E-06
2.38	AGC90367	alanine dehydrogenase	E	K00259	3.39E-04
2.09	AGC89451	carbamate kinase	E	K00926	5.86E-03
1.99	AGC90524	glycine dehydrogenase subunit 1	E	K00282	3.36E-04
1.82	AGC90101	serine hydroxymethyltransferase	E	K00600	2.83E-04

1.81	AGC90587	3-phosphoshikimate 1-carboxyvinyltransferase	E	K00800	9.35E-04
1.69	AGC91147	cysteine desulfurase	E	K11717	2.25E-03
1.58	AGC91234	peptidase T	E	K01258	4.78E-04
4.20	AGC89630	acetolactate synthase	EH	K01652	1.17E-04
1.54	AGC91419	branched-chain amino acid aminotransferase	EH	K00826	2.84E-03
5.84	AGC91651	hypothetical protein	ER	N/A	6.93E-05
4.79	AGC90289	MutT/NUDIX hydrolase family protein	F	N/A	2.29E-06
3.48	AGC90560	MutT/nudix family protein	F	K01515	5.38E-03
2.46	AGC90462	5'-methylthioadenosine/S-adenosylhomocysteine nucleosidase	F	K01243	1.29E-03
2.07	AGC90940	nucleoside-triphosphatase	F	K02428	4.28E-03
1.80	AGC90341	formate--tetrahydrofolate ligase	F	K01938	9.42E-03
1.53	AGC90075	purine nucleoside phosphorylase	F	K03784	3.32E-03
2.32	AGC91483	ribose-phosphate pyrophosphokinase	FE	K00948	5.41E-04
14.01	AGC91608	beta-D-glucuronidase	G	K01195	1.42E-05
9.54	AGC90166	hypothetical protein	G	K00847	9.62E-03
4.70	AGC90387	glyceraldehyde 3-phosphate dehydrogenase 2	G	K00134	1.19E-04
4.35	AGC89614	6-phospho-beta-glucosidase	G	K01223	2.08E-03
3.73	AGC89894	ribose-5-phosphate isomerase A	G	K01807	4.16E-04
3.56	AGC90078	Phosphopentomutase	G	K01839	3.21E-05
3.35	AGC90377	6-phosphofructokinase	G	K00850	1.03E-04
3.12	AGC91031	bifunctional autolysin	G	K13714	3.71E-03
2.84	AGC90980	hypothetical protein	G	N/A	6.86E-05
2.79	AGC91611	glucuronate isomerase	G	K01812	6.63E-03
2.42	AGC91563	phosphoglycerate mutase	G	K01834	2.10E-04
2.22	AGC90553	alpha-D-1,4-glucosidase	G	K01187	1.18E-04
2.14	AGC91617	hypothetical protein	G	K01207	5.46E-05
2.13	AGC89799	Phosphoglyceromutase	G	K01834	5.40E-04
2.06	AGC89611	pyruvate phosphate dikinase	G	K01006	4.16E-03
2.05	AGC90297	putative translaldolase	G	K00616	5.69E-05

1.92	AGC89618	fructose-1,6-bisphosphate aldolase	G	K01623	7.92E-06
1.88	AGC90058	phosphoglucosamine mutase	G	K03431	1.14E-03
1.85	AGC90378	pyruvate kinase	G	K00873	1.79E-03
1.79	AGC91065	inorganic polyphosphate/ATP-NAD kinase	G	K00858	3.03E-04
1.72	AGC90701	Transketolase	G	K00615	9.96E-04
1.67	AGC91331	dihydroxyacetone kinase subunit DhaK	G	K05878	1.24E-04
1.65	AGC89892	aldose 1-epimerase	G	K01785	8.45E-03
7.28	AGC91464	pyridoxal biosynthesis lyase PdxS	H	K06215	2.70E-05
4.21	AGC91045	putative lipoate-protein ligase A	H	K03800	2.60E-03
4.02	AGC90769	riboflavin kinase / FAD synthase ribC	H	K11753	1.84E-03
3.24	AGC91397	phosphomethylpyrimidine kinase	H	K00868	1.78E-04
3.10	AGC90406	glutamate-1-semialdehyde aminotransferase	H	K01845	3.47E-04
2.05	AGC89778	2-dehydropantoate 2-reductase	H	K00077	6.55E-03
1.62	AGC90283	Ferrochelatase	H	K01772	2.11E-03
4.72	AGC90237	cytosolic long-chain acyl-CoA thioester hydrolase family protein	I	N/A	6.63E-05
3.14	AGC91386	mevalonate kinase	I	K00869	1.73E-03
2.58	AGC89675	hydroxymethylglutaryl-CoA synthase	I	K01641	1.72E-03
2.05	AGC90811	putative phosphate acyltransferase	I	K03621	3.12E-04
1.65	AGC89610	acetyl-CoA synthetase	I	K01895	5.68E-04
1.58	AGC89676	hydroxymethylglutaryl-CoA reductase, degradative	I	K00054	4.68E-03
1.57	AGC91061	enoyl-(acyl carrier protein) reductase	I	K00208	1.06E-03
1.55	AGC90810	malonyl CoA-acyl carrier protein transacylase	I	K00645	8.84E-04
2.01	AGC89770	acetoin reductase	IQR	K03366	2.71E-03
1.92	AGC90809	3-oxoacyl-ACP reductase	IQR	K00059	6.73E-04
2.00	AGC90227	methionine aminopeptidase	J	K01265	5.59E-03
1.64	AGC90824	methionyl-tRNA formyltransferase	J	K00604	8.01E-03

1.62	AGC90001	50S ribosomal protein L6	J	K02933	3.60E-03
1.89	AGC91291	MarR family transcriptional regulator	K	N/A	1.88E-04
12.34	AGC90384	DNA polymerase I	L	K02335	4.20E-03
6.14	AGC91495	hypothetical protein	L	K07461	1.04E-04
4.27	AGC91506	DNA polymerase III subunits gamma and tau	L	K02343	3.18E-04
2.82	AGC91492	TatD family deoxyribonuclease	L	K03424	2.02E-05
2.64	AGC90756	recombinase A	L	K03553	9.04E-05
2.33	AGC90689	DNA topoisomerase IV subunit B	L	K02622	2.79E-03
1.97	AGC90271	putative DNA repair exonuclease	L	N/A	8.91E-05
1.79	AGC90364	hypothetical protein	L	K00571	1.39E-03
1.68	AGC90688	DNA topoisomerase IV subunit A	L	K02621	7.64E-04
4.30	AGC91484	bifunctional N-acetylglucosamine-1-phosphate uridyltransferase	M	K04042	7.86E-05
3.27	AGC89738	Glycosyltransferase	M	N/A	3.00E-05
2.00	AGC91519	glycosyl transferase, group 1 family protein	M	K00712	6.46E-03
1.99	AGC91239	UDP-N-acetylenolpyruvoylglucosamine reductase	M	K00075	1.41E-05
1.92	AGC90601	penicillin-binding protein 2	M	K05366	8.57E-03
1.67	AGC89857	teichoic acid biosynthesis protein F	M	N/A	6.22E-04
2.57	AGC89781	putative NAD-dependent epimerase/dehydratase	MG	N/A	8.53E-03
2.18	AGC91420	NAD dependent epimerase/dehydratase family protein	MG	N/A	1.46E-03
2.09	AGC90069	hypothetical protein	MG	N/A	2.38E-04
17.24	AGC91125	hypothetical protein	O	N/A	1.19E-03
4.72	AGC91149	FeS assembly ATPase SufC	O	K09013	1.82E-04
3.83	AGC90480	molecular chaperone DnaK	O	K04043	2.92E-04
2.73	AGC90177	co-chaperonin GroES	O	K04078	1.54E-03
2.46	AGC91162	OsmC/Ohr family protein	O	N/A	3.61E-05
2.44	AGC91472	cell division protein FtsH	O	K03798	1.22E-03

2.26	AGC91458	ATP-dependent Clp protease, ATP-binding subunit ClpC	O	K03696	5.84E-03
2.13	AGC90178	chaperonin GroEL	O	K04077	7.79E-03
2.11	AGC90399	ATP-dependent protease ATP-binding subunit ClpX	O	K03544	1.36E-03
1.91	AGC91084	clpB protein	O	K03695	1.19E-03
1.88	AGC90622	methionine sulfoxide reductase B	O	K07305	4.40E-03
1.88	AGC90574	pyridine nucleotide-disulfide oxidoreductase family protein	O	N/A	2.71E-03
1.58	AGC90621	methionine sulfoxide reductase A	O	K07304	2.46E-03
2.50	AGC91209	ATP-dependent Clp protease proteolytic subunit	OU	K01358	5.12E-04
4.26	AGC91318	hypothetical protein	P	K07220	1.11E-03
1.58	AGC91090	fumarylacetoacetate (FAA) hydrolase family protein	Q	N/A	3.65E-04
6.41	AGC91182	aldo/keto reductase family protein	R	N/A	3.40E-03
4.05	AGC89564	cell wall surface anchor family protein	R	N/A	2.87E-05
3.39	AGC89606	hypothetical protein	R	N/A	3.17E-05
3.07	AGC89856	Acetyltransferase	R	N/A	5.58E-03
2.56	AGC91414	Hydrolase (HAD superfamily) protein	R	K07024	8.74E-03
2.36	AGC90464	GTPase YqeH	R	K06948	7.14E-03
2.32	AGC89862	malate:quinone oxidoreductase	R	K00116	4.39E-04
1.91	AGC91289	aldo keto reductase	R	N/A	6.87E-04
1.91	AGC89405	Proline rich protein	R	K03392	6.52E-03
1.83	AGC89647	Hydrolase	R	N/A	8.72E-04
1.82	AGC89613	malate:quinone oxidoreductase	R	K00116	6.69E-03
1.81	AGC90396	MutT/NUDIX hydrolase family protein	R	N/A	2.28E-03
1.77	AGC90814	hypothetical protein	R	K07030	9.70E-04
1.73	AGC90852	hypothetical protein	R	K06997	4.63E-05
1.68	AGC90608	hypothetical protein	R	N/A	2.79E-03
1.63	AGC90654	ABC transporter ATP-binding protein	R	N/A	6.92E-04

1.61	AGC90681	4-oxalocrotonate tautomerase	R	K01821	6.35E-03
1.56	AGC91207	Cell division inhibitor	R	K07071	3.18E-03
4.73	AGC89533	hypothetical protein	S	N/A	2.79E-05
4.69	AGC90851	YlmF protein	S	K09772	1.98E-04
4.19	AGC89642	hypothetical protein	S	K04750	8.29E-03
3.41	AGC91541	hypothetical protein	S	N/A	1.88E-03
3.31	AGC91253	putative 5'(3')-deoxyribonucleotidase	S	N/A	9.91E-03
3.07	AGC89727	hypothetical protein	S	N/A	6.63E-04
2.59	AGC91460	hypothetical protein	S	N/A	1.82E-05
2.33	AGC90514	hypothetical protein	S	N/A	9.92E-03
2.26	AGC91066	GTP pyrophosphokinase	S	K07816	3.95E-05
2.02	AGC89887	hypothetical protein	S	N/A	2.82E-03
1.93	AGC90815	hypothetical protein	S	N/A	6.59E-03
3.91	AGC90974	GTP-binding protein TypA/BipA	T	K06207	5.19E-04
2.52	AGC89471	serine threonine rich antigen	T	N/A	2.79E-03
1.75	AGC90489	PhoH family protein	T	K06217	1.75E-04
1.63	AGC89450	transcriptional regulator ArcR	T	N/A	3.54E-03
1.56	AGC90366	universal stress protein	T	N/A	4.08E-03
2.13	AGC89805	fmhA protein	V	N/A	3.24E-04
6.83	AGC90754	hypothetical protein	X	N/A	2.50E-03
6.14	AGC91700	hypothetical protein	X	N/A	5.81E-03
4.85	AGC90088	hypothetical protein	X	N/A	5.21E-04
3.35	AGC89384	hypothetical protein	X	N/A	1.47E-03
3.07	AGC90718	hypothetical protein	X	N/A	8.01E-03
2.04	AGC91206	hypothetical protein	X	N/A	1.73E-04
2.02	AGC90636	ABC transporter-like protein	X	N/A	8.01E-03
1.73	AGC90083	hypothetical protein	X	N/A	1.92E-03

Down-regulated proteins

Fold change	Accession number	Protein name (from gene bank)	COG class	KEGG	*p*-value
0.12	AGC89487	pyruvate formate-lyase	C	K00656	9.30E-06
0.21	AGC89860	putative L-lactate permease 2	C	K03303	2.09E-06
0.41	AGC90107	F0F1 ATP synthase subunit B	C	K02109	4.41E-04

0.44	AGC91023	quinol oxidase subunit 1	C	K02827	7.73E-05
0.45	AGC91022	hypothetical protein	C	K02826	5.93E-06
0.55	AGC90112	F0F1 ATP synthase subunit epsilon	C	K02114	3.94E-06
0.55	AGC90108	F0F1 ATP synthase subunit delta	C	K02113	4.08E-03
0.60	AGC90110	F0F1 ATP synthase subunit gamma	C	K02115	2.12E-03
0.63	AGC90545	dihydrolipoamide dehydrogenase	C	K00382	2.42E-03
0.04	AGC89705	D-lactate dehydrogenase	CHR	K03778	2.91E-03
0.31	AGC90256	amino acid ABC transporter substrate-binding protein	E	K02029	8.62E-06
0.42	AGC90714	threonine synthase	E	K01733	3.80E-03
0.60	AGC90715	homoserine dehydrogenase	E	K00003	7.84E-03
0.63	AGC89946	urease subunit gamma	E	K01430	3.87E-04
0.23	AGC89685	zinc-containing alcohol dehydrogenase	ER	K00001	6.48E-06
0.57	AGC89802	ABC transporter periplasmic amino acid-binding protein	ET	K02424	8.32E-06
0.24	AGC91009	bifunctional phosphoribosylaminoimidazolecarboxamide formyltransferase	F	K00602	1.74E-04
0.29	AGC91016	phosphoribosylaminoimidazole-succinocarboxamide synthase	F	K01923	1.70E-03
0.49	AGC91551	xanthine phosphoribosyltransferase	F	K03816	5.40E-06
0.59	AGC90618	thymidylate synthase	F	K00560	3.41E-04
0.62	AGC90208	adenylosuccinate lyase	F	K01756	1.74E-05
0.06	AGC90030	6-phospho-beta-galactosidase	G	K01220	4.23E-03
0.53	AGC91203	glyceraldehyde-3-phosphate dehydrogenase	G	K00134	1.79E-04
0.59	AGC91200	Phosphoglyceromutase	G	K15633	1.70E-05
0.60	AGC91201	triosephosphate isomerase	G	K01803	3.31E-05
0.47	AGC91342	tagG protein, teichoic acid ABC transporter protein	GM	K09692	6.62E-03
0.63	AGC90403	porphobilinogen deaminase	H	K01749	3.06E-03
0.52	AGC89842	Oxidoreductase	IQR	N/A	6.69E-03
0.23	AGC91442	50S ribosomal protein L7/L12	J	K02935	1.39E-03
0.34	AGC90004	50S ribosomal protein L30	J	K02907	1.17E-03

0.39	AGC90002	50S ribosomal protein L18	J	K02881	1.13E-03
0.40	AGC90009	50S ribosomal protein L36	J	K02919	8.77E-03
0.44	AGC90816	50S ribosomal protein L28	J	K02902	7.14E-04
0.45	AGC90509	50S ribosomal protein L33	J	K02913	6.02E-03
0.46	AGC89991	50S ribosomal protein L22	J	K02890	2.26E-04
0.46	AGC90395	50S ribosomal protein L20	J	K02887	3.05E-03
0.46	AGC90000	30S ribosomal protein S8	J	K02994	3.88E-04
0.49	AGC89993	50S ribosomal protein L16	J	K02878	1.91E-03
0.49	AGC90799	50S ribosomal protein L19	J	K02884	2.15E-03
0.49	AGC91588	30S ribosomal protein S18	J	K02963	1.49E-04
0.50	AGC90018	50S ribosomal protein L13	J	K02871	4.51E-04
0.51	AGC91590	30S ribosomal protein S6	J	K02990	6.32E-05
0.53	AGC89989	50S ribosomal protein L2	J	K02886	5.35E-05
0.54	AGC90013	50S ribosomal protein L17	J	K02879	1.62E-03
0.56	AGC90801	16S rRNA-processing protein RimM	J	K02860	9.78E-04
0.56	AGC91223	peptide chain release factor 2	J	K02836	4.60E-03
0.56	AGC89985	30S ribosomal protein S10	J	K02946	1.46E-04
0.56	AGC90005	50S ribosomal protein L15	J	K02876	1.62E-03
0.57	AGC91475	hypothetical protein	J	K07571	3.92E-05
0.57	AGC90427	queuine tRNA-ribosyltransferase	J	K00773	2.63E-03
0.57	AGC90597	asparaginyl-tRNA ligase	J	K01893	5.59E-05
0.59	AGC90784	elongation factor Ts	J	K02357	4.06E-04
0.59	AGC89994	50S ribosomal protein L29	J	K02904	6.13E-05
0.60	AGC89997	50S ribosomal protein L24	J	K02895	4.43E-03
0.60	AGC91444	50S ribosomal protein L1	J	K02863	3.04E-05
0.61	AGC89988	50S ribosomal protein L23	J	K02892	5.75E-04
0.62	AGC91436	30S ribosomal protein S7	J	K02992	1.45E-04
0.62	AGC91443	50S ribosomal protein L10	J	K02864	1.96E-04
0.63	AGC89995	30S ribosomal protein S17	J	K02961	6.05E-04
0.65	AGC90847	isoleucyl-tRNA ligase	J	K01870	1.00E-04
0.66	AGC90010	30S ribosomal protein S13	J	K02952	7.44E-04
0.23	AGC90086	DNA-directed RNA polymerase subunit delta	K	K03048	2.67E-06
0.30	AGC91440	DNA-directed RNA polymerase	K	K03043	3.65E-04

		subunit beta			
0.40	AGC90120	TenA family transcription regulator	K	K03707	5.08E-04
0.40	AGC89836	TetR family regulatory protein	K	N/A	1.76E-03
0.41	AGC90143	RNA polymerase sigma factor SigB	K	K03090	7.26E-03
0.44	AGC90455	transcription elongation factor GreA	K	K03624	2.24E-05
0.48	AGC89430	putative transcriptional regulator	K	N/A	1.56E-03
0.57	AGC89826	MarR family transcriptional regulator	K	N/A	9.66E-04
0.42	AGC90696	hypothetical protein	L	K03546	2.53E-04
0.53	AGC91410	hypothetical protein	M	N/A	2.64E-03
0.58	AGC90593	putative glycosyltransferase	M	N/A	7.93E-03
0.66	AGC91106	peptidyl-prolyl cis-trans isomerase	O	K03768	2.23E-03
0.17	AGC90043	hypothetical protein	P	K02016	1.12E-06
0.20	AGC91152	hypothetical protein	P	K02073	1.41E-05
0.32	AGC91154	ABC transporter ATP-binding protein	P	K02071	7.11E-04
0.38	AGC89958	molybdate ABC transporter periplasmic molybdate-binding protein	P	K02020	1.00E-03
0.40	AGC89664	putative heavy-metal-associated protein	P	K07213	6.22E-04
0.44	AGC91349	ABC transporter substrate-binding protein	P	K09818	1.76E-03
0.34	AGC90039	putative siderophore biosynthesis protein	Q	N/A	5.64E-05
0.37	AGC90081	Amidase	R	N/A	2.95E-03
0.40	AGC90422	GTPase CgtA	R	K03979	1.07E-03
0.45	AGC90515	metallo-beta-lactamase superfamily protein	R	N/A	1.31E-05
0.47	AGC90766	Ribonuclease	R	K12574	4.65E-06
0.63	AGC90762	M16 family peptidase	R	N/A	6.39E-03
0.08	AGC90684	FmtC (MrpF) protein	S	K14205	8.93E-03
0.08	AGC90497	PEP synthetase regulatory protein	S	K09773	9.06E-03
0.11	AGC89813	NirR family transcriptional	S	N/A	3.44E-07

		regulator			
0.44	AGC90993	hypothetical protein	S	N/A	1.10E-04
0.60	AGC91229	hypothetical protein	S	N/A	7.11E-04
0.50	AGC90140	sigmaB regulation protein	TK	K07315	4.31E-03
0.50	AGC89350	two-component response regulator SA14-24	TK	K07668	2.26E-04
0.55	AGC90803	signal recognition particle protein	U	K03106	5.62E-04
0.43	AGC91337	ABC transporter permease/ATP-binding protein	V	K06147	3.47E-04
0.47	AGC90669	femB protein	V	K11695	1.09E-04
0.52	AGC90670	methicillin resistance protein FemA	V	K11694	2.77E-04
0.08	AGC91399	cysteine protease precursor SspB	X	K08258	2.53E-04
0.10	AGC90867	phenol soluble modulin beta 1	X	N/A	3.08E-05
0.11	AGC90866	antibacterial protein	X	N/A	4.39E-06
0.41	AGC89549	cysteine protease precursor	X	K13715	2.86E-03
0.45	AGC91164	Acetyltransferase	X	N/A	1.18E-05
0.46	AGC91312	hypothetical protein	X	N/A	1.24E-05
0.46	AGC90190	hypothetical protein	X	N/A	7.96E-03
0.53	AGC90048	hypothetical protein	X	N/A	1.03E-04

3.3.1.1 Metabolic pathways of SG1 involved in butanol adaption

The quantitative proteomic analysis of *Staphylococcus warneri* SG1 enables us to identify important metabolic pathways that are regulated by SG1 upon butanol challenge. Based on the KEGG pathway database, we reconstructed a metabolic pathway map of SG1 for butanol adaption including pathways of glycolysis, Krebs cycle, pentose phosphate pathway, fatty acid biosynthesis, amino sugar and nucleotide sugar metabolism, peptidoglycan biosynthesis and oxidative phosphorylation (Figure 3.5). Proteins that have chaperone functions, responsible for oxidative stress response or glucuronate interconversion were also summarized. The reconstructed pathway map (Figure 3.5) depicts the dynamic responses of *Staphylococcus warneri* SG1 upon exposure to 1.5 % 1-Butanol and illustrates a complicated mechanism for butanol adaptation.

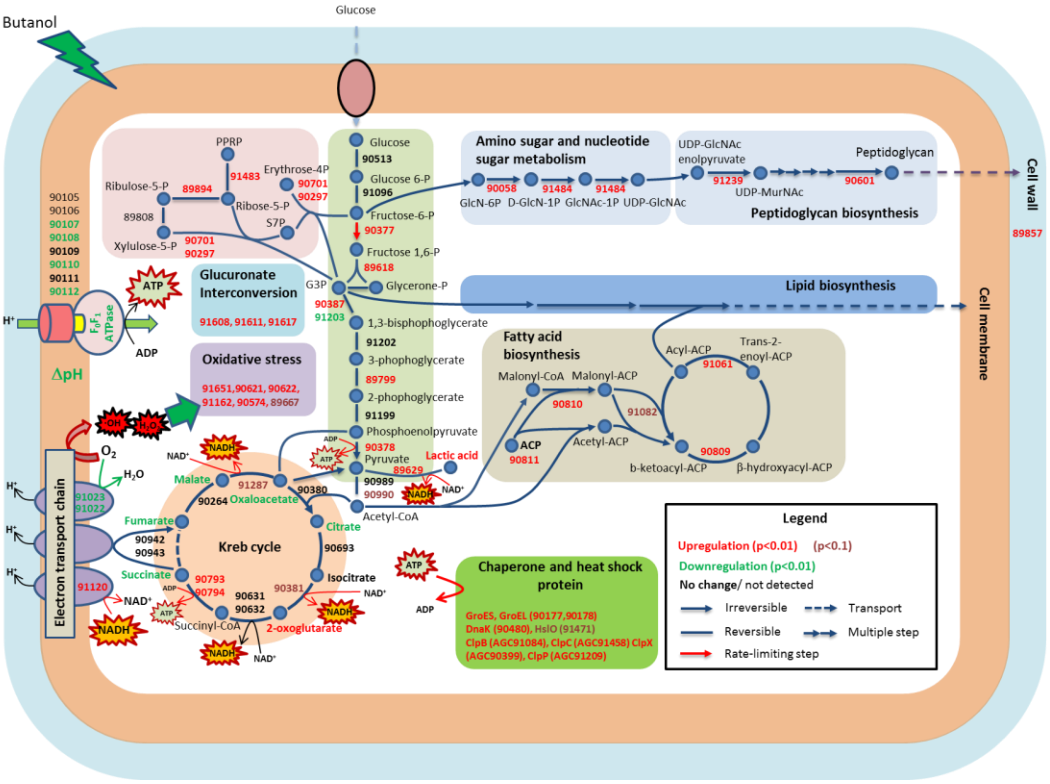


Figure 3.5 Overall schematic representation of the dynamic responses in *Staph. warneri* SG1 upon exposure to 1.5 % 1-Butanol. Accession numbers for proteins which were up-regulated at least 1.5-fold (p-value < 0.01) are marked in red, whereas those corresponding to proteins which were down-regulated by at least 0.67-fold are marked in green. Metabolites are represented by circles while ATP and NADH are also highlighted.

3.3.1.2 Membrane and membrane composition

The first line of defense against solvent exposure is the cell envelope/membrane. An unusual extracellular capsule for *Staphylococcus warneri* ZZ1 was developed upon toluene exposure[8]. In our studies, we observed that $BtOH^+$ SG1 cells were resistant to chemical lysis by lysostaphin. Since lysostaphin is a glycylglycine endopeptidase that hydrolyzes the cross-bridges in the peptidoglycan layers of Staphyloccocci[34,35], we can infer that butanol exposure resulted in alteration of the peptidoglycan structure/composition in SG1. In our

2D-LC-MS/MS studies, we observed a >1.5-fold increase in the abundances of UDP-N-acetylglucosamine diphosphorylase (AGC91484), UDP-N-acetylenolpyruvoylglucosamine reductase (AGC91239), phosphoglucosamine mutase (AGC90058), epimerases (AGC89781, AGC91420 and AGC90069), glycosyltransferases (AGC89738 and AGC91519), penicillin-binding protein 2 (AGC90601) and teichoic acid biosynthesis protein F (AGC89857) (Table 3.1). These proteins are crucial for the biosynthesis of the peptidoglycan precursors N-acetylglucosamine and N-acetylmuramic acid, as well as catalyzing their cross-linking via a short polypeptide[36,37]. Meanwhile, the increased amount of teichoic acid synthesized would allow additional modification of the peptidoglycan layer to strengthen it[37,38]. Altogether, the increased expression of cell wall synthesis/modification enzymes suggests an elaborate adaptive mechanism for SG1 to survive in the presence of butanol.

Changes in fatty acid and phospholipid compositions upon solvent exposure are well documented for both Gram-negative and Gram-positive bacteria. However, the reported effects were contradictory. Studies done on *E. coli* and *Pseudomonas putida* demonstrated denser membrane packing and therefore a decrease in membrane fluidity upon organic solvent exposure[6]. In contrast, *Staphylococcus haemolyticus* and *Bacillus sp*. strain ORAs2 adapted to solvent challenge by increasing membrane fluidity[7,39]. In this study, we found a branched-chain amino acid transaminase (AGC91419) was up-regulated 1.5-fold. This protein initiates the first step of branched fatty acids synthesis by converting isoleucine, leucine and valine to the corresponding α-keto acids. Thus it appears that SG1 is increasing the membrane fluidity in order to combat butanol stress. Several other enzymes involved in the fatty acid synthesis machinery were also up-regulated in the presence of butanol. Malonyl CoA-acyl carrier protein transacylase (AGC90810), 3-oxoacyl-[acyl-carrier protein] reductase (AGC90809) and enoyl-[acyl-carrier protein] reductase I (FAGC91061) were all up-regulated, indicating enhanced production of fatty acid in SG1 under butanol stress. Finally,

we observed a 2.1-fold up-regulation of the fatty acid/phospholipid synthesis protein PlsX (AGC90811), which has been shown to catalyze the synthesis of fatty acyl-phosphate[40] and postulated to play a role in fatty acid metabolism by regulating the intracellular concentration of acyl-[acyl carrier protein][41].The *plsX* gene is part of a locus that comprise genes encoding 3-oxoacyl-[acyl-carrier protein] reductase and enoyl-[acyl-carrier protein] reductase mentioned above, as well as a gene encoding for a fatty acid biosynthesis transcriptional regulator (AGC90812). Altogether, it appears that up-regulation of fatty acid synthesis enzymes under the $BtOH^+$ condition allows SG1 to overcome butanol stress by increasing fatty acid synthesis, especially those which contain branched chains to increase the overall membrane fluidity.

3.3.1.3 Changes in energy metabolism

A key mechanism of solvent toxicity is the partition of the solvent into the cytoplasmic membrane, causing leakage of ions and small molecules across the lipid bilayer[42-44]. The loss of the proton gradient and membrane potential leads to deprivation in ATP synthesis and inhibition of transport processes that are coupled to the proton motive force. In SG1, the enzymes involved in oxidative phosphorylation are differentially regulated upon butanol challenge: the soluble NADH dehydrogenase (AGC91120) was up-regulated > 5-fold; expression of the succinate dehydrogenase complex was unchanged; the terminal cytochrome *aa*3 oxidase (AGC91022, subunit I; AGC91023, subunit II) and ATP synthase (AGC90107, B subunit of F_o; AGC90108, δ subunit of F_1; AGC90110, γ subunit of F_1; AGC90112, ε subunit of F_1) were down-regulated ~ 2-fold. Different studies in the literature show that ATP synthase can either be down-regulated[21,45] or up-regulated[23,46] upon solvent stress.

To compensate for the decreased level of energy production from oxidative phosphorylation, cells must utilize other methods of ATP generation, such as substrate level phosphorylation. In our 2-MEGA labeling study on SG1, the

majority of proteins that are involved in glycolysis and the tricarboxylic acid cycle were either unchanged or up-regulated upon butanol challenge (Figure 3.5, COG class G). Interestingly, we observed the key reactions of these pathways to be highly influenced by the presence of butanol. In glycolysis, the irreversible conversion of β-D-fructose 6-phosphate to β-D-fructose 1,6-bisphosphate represents a crucial rate-limiting step. Phosphofructokinase (AGC90377), the enzyme that catalyzes this reaction, was up-regulated 3.4-fold when cells were grown in the presence of butanol. The final enzyme in glycolysis, pyruvate kinase (AGC90378), generates ATP at the substrate level and was up-regulated 1.9-fold under the $BtOH^+$ condition. In the tricarboxylic acid cycle, succinyl-CoA synthetase (α subunit, AGC90793; β subunit, AGC90794) can also generate either ATP or GTP, and it was up-regulated 3.8-fold. Enzymes that are involved in NADH generation, such as glyceraldehyde 3-phosphate dehydrogenase (AGC90387) and malate dehydrogenase (AGC90374), were also up-regulated (Figure 3.5). The increased expression of key proteins involved in the central carbohydrate metabolic pathways is undoubtedly a cellular response to compensate for the decreased ATP synthesis from oxidative phosphorylation and to meet the high energy demands of combatting butanol stress[23,45,46].

Comparative genomics between *Staphylococcus warneri* SG1 and other *Staphylococci* species suggested a number of gene products which may contribute to the solvent tolerance properties of SG1[11]. One distinction is the presence of a gene cluster that encodes for enzymes involved in glucuronate interconversion (AGC91608 - AGC91617). 2-MEGA labeling showed that β-D-glucuronidase (AGC91608), glucuronate isomerase (AGC91611) and beta-N-acetylhexosaminidase (AGC91617) were induced 14.0-fold, 2.8-fold, 2.1-fold, respectively and none of the enzymes in this locus involved in glucuronate interconversion were down-regulated upon butanol exposure. To our knowledge, this is the first report of a positive correlation between glucuronate interconversion and solvent tolerance and their mechanistic relationship is unclear.

One possibility is that these unusual sugars are used to synthesize a hydrophilic extracellular capsule to repel organic solvents, as exemplified by *Staphylococcus warneri* ZZ1 upon toluene exposure[8,47].

3.3.1.4 Global stress responses

Exposure to organic solvents typically triggers a global cellular response involving heat shock proteins, oxidative stress proteins, and transcriptional activators/repressors[6,47,48]. The 2-MEGA labeling studies showed that butanol exposure to SG1 resulted in COG class O (post-translational modification, protein turnover, chaperones) being the second most elevated COG class only behind carbohydrate transport and metabolism (Figure 3.3). Interestingly, the most up-regulated proteins in COG class O upon butanol challenge were a hypothetical protein with a NifU-like domain (AGC91127) and an iron-sulfur cluster assembly protein SufC (AGC91149), which were increased 17.2 and 4.7-fold, respectively. In plants and bacteria, the NifU protein acts as a scaffold for iron-sulfur cluster formation[49,50]. The up-regulation of a NifU-like protein along with SufC suggests that iron-sulfur cluster proteins in SG1 were damaged in the presence of butanol. If this is true, then labile iron from damaged proteins could lead to increased oxidative stress in the cell and trigger repair proteins to counteract oxidative damage.

A number of proteins belonging to the Clp superfamily of proteases were induced at least 1.5-fold in the $BtOH^+$ condition (Table 3.1: ClpC, AGC91458; ClpX, AGC90399; ClpB, AGC91084; ClpP, AGC91209). In *S. aureus*, the proteolytic activity of ClpP is critical for pathogenicity, stress response, metal homeostasis, prevention of autolysis, and activation of the heat shock regulon[51,52]. The Clp proteases are also activated upon solvent stress in the strict anaerobe *C. acetobutylicum*[19,20,53]. In agreement with those similar studies, several other molecular chaperones, notably DnaK (AGC90480), GroES (AGC90177) and GroEL (AGC90178), were also induced in SG1 in the presence of butanol.

In addition to the chaperones and Clp proteases, we also observed increased abundances of two Zn-containing alcohol dehydrogenases (AGC90034 and AGC91651) and an aldehyde dehydrogenase (AGC90092) that could potentially help combat oxidative stress. Interestingly, we also saw up-regulation of methionine-(S)-sulfoxide reductase (MsrA, AGC90621) and methionine-(R)-sulfoxide reductase (MsrB, AGC90622) by 1.6- and 1.9-fold, respectively, in the $BtOH^+$ condition. As shown in studies on other bacterial models, MsrA, as well as MsrB to a lesser extent, are crucial for the amelioration of cellular damage from reactive oxygen species[54,55]. Finally, peroxiredoxin (OsmC, AGC91162) was also up-regulated 2.5-fold in SG1 upon butanol challenge. This ubiquitous redox sensitive protein is a good biomarker for oxidative stress[56,57]. The combination of proteases, molecular chaperones, redox sensors, and scavengers of reactive oxygen species which are up-regulated paints a strong picture of oxidative stress induction in SG1 in the presence of butanol.

3.3.2 Metabolomic study on amine- containing metabolites and TCA cycle metabolites

In our metabolomics study, we applied a differential isotopic labeling technique to target the amine-, phenol- and acid-containing sub-metabolomes. The comparison between abundances of $BtOH^-$ and $BtOH^+$ metabolites was based on the ratio of ^{12}C-labeled individual samples to the ^{13}C-labeled pooled reference sample[58,59]. This labeling technique allows signal enhancement of 10- to 1000-fold and is particularly useful for monitoring amino acids, polyamines and organic acids in biological samples. As we will see below, these compounds play an important role in response to butanol stress and a correlation with the proteomics results can be observed.

3.3.2.1 Analysis of amine- and phenol-containing metabolites.

On average, 605 ± 86 peak pairs were extracted from triplicate $BtOH^-$ and $BtOH^+$ samples. Principal component analysis was first applied to evaluate

whether the addition of butanol had an effect on the metabolite profiles of SG1. As shown in Figure 3.6, there was a clear distinction between $BtOH^-$ and $BtOH^+$, and the separation is mainly reflected by the first principal component, indicating that the presence of butanol lead to a significant alteration of the metabolite levels in SG1. Using a volcano plot to find individual metabolites which varied by at least 1.5-fold with a *p*-value of < 0.01, we identified 94 metabolites, amongst which 15 were unambiguously identified (Table 3.2).

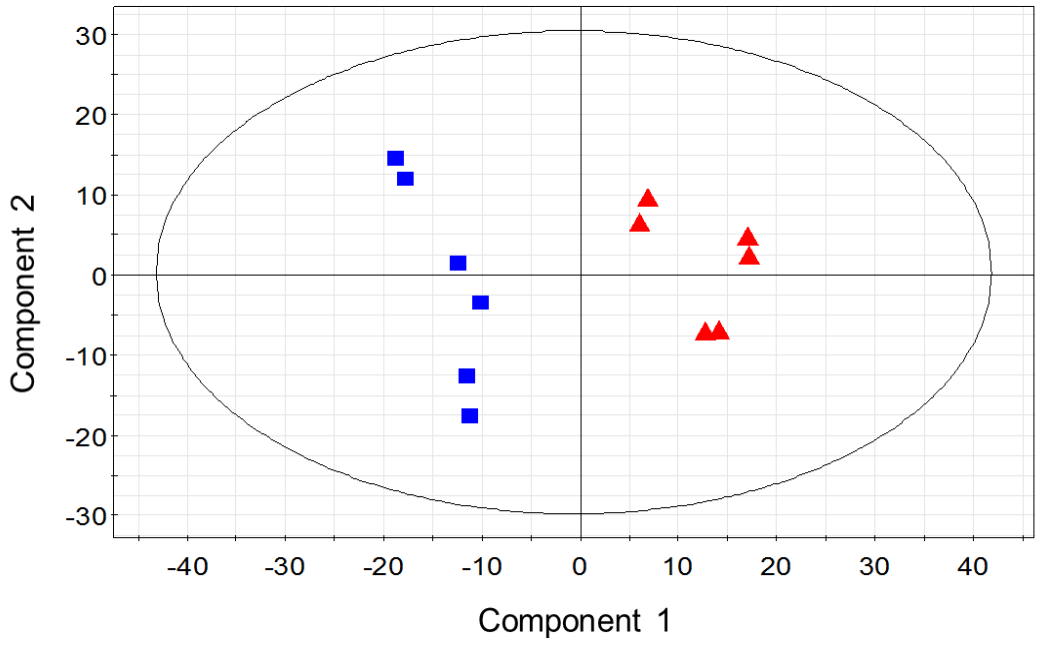


Figure 3.6 Principal component analysis plot of dansylation labeled metabolites from *Staph. warneri* SG1 grown in $BtOH^-$ or $BtOH^+$ medium. The data was from technical duplicate experiments of biological triplicate samples of $BtOH^-$ and $BtOH^+$ lysates. Red triangle and blue boxes represent $BtOH^-$ and $BtOH^+$ cultures, respectively.

Table 3.2 List of differentially expressed amine-containing metabolites that were unambiguously identified. The fold change represents the levels of metabolite seen in $BtOH^+$ versus $BtOH^-$ metabolome.

Metabolite name	Fold change	p-value	avg. rt (mins)	avg. m/z (light)	avg. m/z (heavy)
N1-Acetylspermine	21.64	7.17E-08	25.1	315.4672	317.4742
Spermine	6.76	1.65E-05	28.9	379.1478	381.8232
Cadaverine	3.99	3.83E-03	23.6	285.1158	287.1226
Arginine	2.21	1.46E-04	4.4	408.1703	410.1769
Pyrrolidinone	1.97	2.80E-04	15.7	319.1115	321.1181
γ-aminobutyric acid	1.89	6.16E-04	9.9	337.1220	339.1287
Threonine	1.70	1.32E-03	8.0	353.1172	355.1239
Taurine	0.61	2.71E-03	3.9	359.0734	361.0800
Methionine sulfoxide	0.53	6.98E-03	6.8	399.1046	401.1115
Glucosamine	0.47	7.94E-05	4.4	413.1383	415.1450
Lysine	0.35	6.43E-04	19.8	307.1115	309.1183
Spermidine	0.30	1.65E-06	26.1	423.1640	426.1741
Ornithine	0.30	3.08E-04	18.7	300.1014	302.1082
Methionine	0.21	9.02E-04	13.3	383.1099	385.1166

As discussed above, butanol stress is linked to the oxidative stress response. We detected higher cadaverine:lysine (Figure 3.7A) and putrescine:ornithine ratios (Figure 3.7B), which is consistent with the up-regulation of Orn/Lys/Arg decarboxylase (AGC91502) in our proteomics data. Cadaverine is a scavenger of superoxide radicals during oxidative stress[60]. Spermine, a polyamine which can protect DNA from free radical attack[61] and inhibit autolysis of gram-negative marine pseudomonad NCMB 845[62,63], was also up-regulated in $BtOH^+$ SG1. Interestingly, *Staphylococcus warneri* SG1 is not predicted to encode a spermine synthase protein, and the physiological function of spermine in bacteria has not been elucidated[64] . A negative aspect of spermine accumulation is the cytotoxicity and inhibition of cell growth[65]. However, the enzyme spermine/spermidine acetyl-transferase can minimize the cytotoxicity of spermine by converting it into N-acetyl-spermine[65], and a similar enzyme (AGC91503) is present in the SG1 genome. Although we only detected the presence of AGC91503 in proteome

profiling and fail to quantify it in the 2MEGA labeling, the amount of N-acetylspermine observed was increased considerably in $BtOH^+$ sample (Figure 3.7C), which suggests the enhanced activity of AGC91503 in the presence of butanol.

The conversion of methionine to methionine sulfoxide is another commonly observed reaction under oxidative stress conditions, wherein, methionine can react with reactive oxygen species via a 2-electron-dependent mechanism to produce methionine sulfoxide[66]. Our metabolomics data show that the ratio of methionine sulfoxide:methionine in $BtOH^+$ increased more than 2.5-fold compared to that in $BtOH^-$ (Figure 3.7D), indicating an increased level of oxidation in response to the oxidative stress caused by butanol. The up-regulation of MsrA and MsrB (discussed above) would allow rapid regeneration of methionine from methionine sulfoxide, allowing continuous disposition of reactive oxygen species.

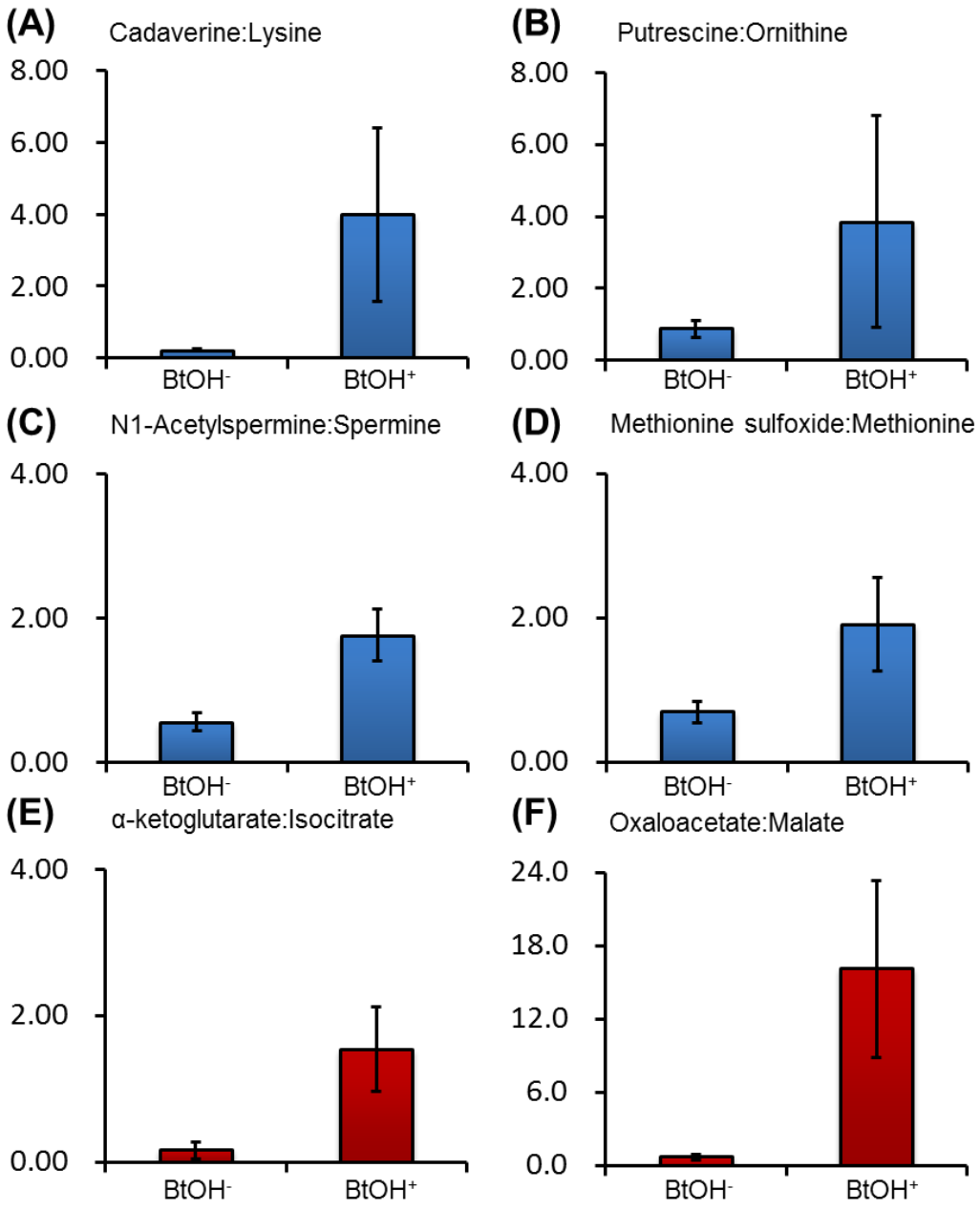


Figure 3.7 Column plots highlighting product:substrate ratios which were increased upon 1-Butanol challenge. The data was from technical duplicate experiments on triplicate $BtOH^-$ and $BtOH^+$ grown cultures. Error bar represents standard deviation of data from technical duplicate experiments on biological triplicate samples.

3.3.2.2 Analysis of TCA cycle carboxylic acids

Prior to isotopic labeling, an extraction step was carried out to remove amine-containing compounds in order to increase the specificity of the reaction for carboxylic acids. With the optimized sample loading, an average of 449 ± 66 peak pairs were extracted from triplicate $BtOH^-$ and $BtOH^+$ samples. In this study, we focused on obtaining dynamic changes of seven TCA cycle carboxylic acids in the presence of butanol (Table 3.3), as the TCA cycle is a critical energy metabolic pathway and the changes in metabolite levels in TCA cycle reflects the regulation of energy metabolism of SG1 under butanol stress.

Table 3.3 Levels of carboxylic acid-containing metabolites in the TCA cycle. Values were derived from dividing ^{12}C-labeled individual samples by ^{13}C-labeled pooled samples. Data from replicate experiments on biological triplicates of *Staph. warneri* cultured in $BtOH^-$ and $BtOH^+$ media were analyzed.

	$BtOH^-$	$BtOH^+$
Citrate	1.83±0.57	1.05±0.53
Isocitrate	1.32±0.58	1.34±0.47
α-ketoglutarate	0.25±0.27	1.87±0.30
Succinate	1.10±0.62	0.52±0.19
Fumarate	1.70±0.73	0.07±0.02
Malate	1.96±0.75	0.05±0.01
Oxaloacetate	1.40±0.75	0.76±0.44

From the 2D-LC-MS/MS proteomic studies, we observed up-regulation of enzymes responsible for critical energy-producing reactions in the TCA cycle. In agreement, the acid profiling results show that the α-ketoglutarate:isocitrate (Figure 3.7E) and oxaloacetate:malate ratios (Figure 3.7F) were increased significantly under butanol challenge. The isocitrate dehydrogenase (isocitrate → α-ketoglutarate) and malate dehydrogenase (malate → oxaloacetate) reactions both produce NADH, which is fed into the oxidative phosphorylation pathway to produce ATP. Interestingly, the other reaction that generates NADH in the TCA

cycle, the α-ketoglutarate dehydrogenase reaction (α-ketoglutarate → succinyl-CoA), was not up-regulated as indicated by the high level of α-ketoglutarate in $BtOH^+$ cells. This could be explained by considering that the enzyme is sensitive to oxidative stress and can be inhibited under such conditions[67]. Since α-ketoglutarate itself can act as an effective scavenger of reactive oxygen species, the stress condition should favor an increased accumulation of α-ketoglutarate[68]. Looking at the downstream intermediates, we saw decreased levels of succinate, fumarate, and malate. Our results indicate that the TCA cycle plays a key role in regulation of the cell response to butanol stress, as it not only promotes the production of energy, but also helps to modulate the oxidative stress condition induced by butanol.

3.4 Conclusion

Understanding the complex interplay that results in solvent tolerance can shed light on how to genetically engineer bacteria for biodegradation, biofuel production and biocatalysis. Our comprehensive proteomic and metabolomic study on SG1 reveals a complicated mechanism for butanol adaptation that spans multiple clusters of orthologous groups. Upon butanol challenge, the structure and composition of cell wall/membrane were altered. In addition, expression of membrane proteins was suppressed. We also saw a strong link between solvent stress and oxidative stress. Many stress response proteins such as chaperones and proteases were up-regulated upon butanol challenge. Key enzymes in carbohydrate metabolism, including those involved in glucuronate interconversions, were also up-regulated to counteract the drop in ATP synthesis via oxidative phosphorylation. Consistent with the proteomic study, interrogation of amine- and phenol- containing metabolites provided strong evidence of oxidative stress when SG1 was exposed to butanol. Analysis of TCA cycle metabolites further confirmed these observations and also indicated the key role of TCA cycle intermediates in the mechanism of butanol tolerance of SG1.

Altogether, the results greatly enrich our knowledge of the mechanisms employed by SG1 to combat butanol toxicity.

3.5 Literature cited

(1) Sikkema, J.; de Bont, J. a.; Poolman, B. *Microbiological Reviews* **1995**, *59*, 201.

(2) Isken, S.; de Bont, J. A. M. *Extremophiles* **1998**, *2*, 229.

(3) Sardessai, Y.; Bhosle, S. *Research in Microbiology* **2002**, *153*, 263.

(4) Ramos, J. L.; Duque, E.; Gallegos, M. T.; Godoy, P.; Ramos-Gonzalez, M. I.; Rojas, A.; Teran, W.; Segura, A. *Annual Review of Microbiology* **2002**, *56*, 743.

(5) Nicolaou, S. A.; Gaida, S. M.; Papoutsakis, E. T. *Metabolic Engineering* **2010**, *12*, 307.

(6) Segura, A.; Molina, L.; Fillet, S.; Krell, T.; Bernal, P.; Munoz-Rojas, J.; Ramos, J. L. *Current Opinion in Biotechnology* **2012**, *23*, 415.

(7) Nielsen, L. E.; Kadavy, D. R.; Rajagopal, S.; Drijber, R.; Nickerson, K. W. *Applied and Environmental Microbiology* **2005**, *71*, 5171.

(8) Zahir, Z.; Seed, K. D.; Dennis, J. J. *Extremophiles* **2006**, *10*, 129.

(9) Grandvalet, C.; Assad-Garcia, J. S.; Chu-Ky, S.; Tollot, M.; Guzzo, J.; Gresti, J.; Tourdot-Marechal, R. *Microbiology* **2008**, *154*, 2611.

(10) Gao, Y.; Dai, J.; Peng, H.; Liu, Y.; Xu, T. *Journal of Applied Microbiology* **2011**, *110*, 472.

(11) Cheng, V. W.; Zhang, G.; Oyedotun, K. S.; Ridgway, D.; Ellison, M. J.; Weiner, J. H. *Genome Announcements* **2013**, *1*, e0003813.

(12) Jones, D.; Woods, D. *Microbiological Reviews* **1986**, *50*, 484

(13) Bermejo, L. L.; Welker, N. E.; Papoutsakis, E. T. *Applied and Environmental Microbiology* **1998**, *64*, 1079.

(14) Steen, E.; Chan, R.; Prasad, N.; Myers, S.; Petzold, C.; Redding, A.; Ouellet, M.; Keasling, J. *Microbial Cell Factories* **2008**, *7*, 36.

(15) Atsumi, S.; Cann, A.; Connor, M.; Shen, C.; Smith, K.; Brynildsen, M.; Chou, K.; Hanai, T.; Liao, J. *Metabolic Engineering* **2008**, *10*, 305

(16) Atsumi, S.; Hanai, T.; Liao, J. C. *Nature* **2008**, *451*, 86.

(17) Tomas, C. A.; Beamish, J.; Papoutsakis, E. T. *Journal of Bacteriology* **2004**, *186*, 2006.

(18) Brynildsen, M. P.; Liao, J. C. *Molecular Systems Biology* **2009**, *5*, 277.

(19) Alsaker, K. V.; Paredes, C.; Papoutsakis, E. T. *Biotechnology and Bioengineering* **2010**, *105*, 1131.

(20) Mao, S.; Luo, Y.; Zhang, T.; Li, J.; Bao, G.; Zhu, Y.; Chen, Z.; Zhang, Y.; Li, Y.; Ma, Y. *Journal of Proteome Research* **2010**, *9*, 3046.

(21) Mao, S.; Luo, Y.; Bao, G.; Zhang, Y.; Li, Y.; Ma, Y. *Molecular BioSystems* **2011**, *7*, 1660.

(22) Segura, A.; Godoy, P.; van Dillewijn, P.; Hurtado, A.; Arroyo, N.; Santacruz, S.; Ramos, J. L. *Journal of Bacteriology* **2005**, *187*, 5937.

(23) Wijte, D.; van Baar, B. L.; Heck, A. J.; Altelaar, A. F. *Journal of Proteome Research* **2011**, *10*, 394.

(24) Ye, X.; Li, L. *Analytical Chemistry* **2012**, *84*, 6181.

(25) Wang, N.; Xie, C.; Young, J. B.; Li, L. *Analytical Chemistry* **2009**, *81*, 1049.

(26) Ji, C.; Zhang, N.; Damaraju, S.; Damaraju, V. L.; Carpenter, P.; Cass, C. E.; Li, L. *Analytica Chimica Acta* **2007**, *585*, 219.

(27) Wang, N.; Li, L. *Analytical Chemistry* **2008**, *80*, 4696.

(28)Tian, X.; Chen, L.; Wang, J.; Qiao, J.; Zhang, W. *Journal of Proteomics* **2013**, *78*, 326.

(29) Hilger, M.; Mann, M. *Journal of Proteome Research* **2012**, *11*, 982.

(30) Tatusov, R. L.; Galperin, M. Y.; Natale, D. A.; Koonin, E. V. *Nucleic Acids Research* **2000**, *28*, 33.

(31) Ogata, H.; Goto, S.; Sato, K.; Fujibuchi, W.; Bono, H.; Kanehisa, M. *Nucleic Acids Research* **1999**, *27*, 29.

(32) Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. *Analytical Chemistry* **2006**, *78*, 779.

(33) Guo, K.; Peng, J.; Zhou, R.; Li, L. *Journal of Chromatography A* **2011**, *1218*, 3689.

(34) Schindler, C. A.; Schuhardt, V. T. *Proceedings of the National Academy of Sciences of the United States of America* **1964**, *51*, 414.

(35) Browder, H. P.; Zygmunt, W. A.; Young, J. R.; Tavormina, P. A. *Biochemical and Biophysical Research Communications* **1965**, *19*, 383.

(36) Silhavy, T. J.; Kahne, D.; Walker, S. *Cold Spring Harbor Perspectives in Biology* **2010**, *2*, a000414.

(37) Reith, J.; Mayer, C. *Applied Microbiology and Biotechnology* **2011**, *92*, 1.

(38) Weidenmaier, C.; Peschel, A. *Nature Reviews Microbiology* **2008**, *6*, 276.

(39) Pepi, M.; Heipieper, H. J.; Fischer, J.; Ruta, M.; Volterrani, M.; Focardi, S. E. *Extremophiles* **2008**, *12*, 343.

(40) Lu, Y. J.; Zhang, Y. M.; Grimes, K. D.; Qi, J.; Lee, R. E.; Rock, C. O. *Molecular Cell* **2006**, *23*, 765.

(41) Yoshimura, M.; Oshima, T.; Ogasawara, N. *BMC Microbiology* **2007**, *7*, 69.

(42) Cartwright, C. P.; juroszek, J.-R.; Beavan, M. J.; Ruby, F. M. S.; De Morais, S. M. F.; Rose, A. H. *Journal of General Microbiology* **1986**, *132*, 369.

(43) Leao, C.; Van Uden, N. *Biochimica et Biophysica Acta* **1984**, *774*, 43.

(44) Bowles, L. K.; Ellefson, W. L. *Applied and Environmental Microbiology* **1985**, *50*, 1165.

(45) Volkers, R. J.; de Jong, A. L.; Hulst, A. G.; van Baar, B. L.; de Bont, J. A.; Wery, J. *Environmental Microbiology* **2006**, *8*, 1674.

(46) Rutherford, B. J.; Dahl, R. H.; Price, R. E.; Szmidt, H. L.; Benke, P. I.; Mukhopadhyay, A.; Keasling, J. D. *Applied and Environmental Microbiology* **2010**, *76*, 1935.

(47) Torres, S.; Pandey, A.; Castro, G. R. *Biotechnology Advances* **2011**, *29*, 442.

(48)Nicolaou, S. a.; Gaida, S. M.; Papoutsakis, E. T. *Metabolic Engineering* **2010**, *12*, 307.

(49) Yuvaniyama, P.; Agar, J. N.; Cash, V. L.; Johnson, M. K.; Dean, D. R. *Proceedings of the National Academy of Sciences of the United States of America* **2000**, *97*, 599.

(50) Smith, A. D.; Jameson, G. N.; Dos Santos, P. C.; Agar, J. N.; Naik, S.; Krebs, C.; Frazzon, J.; Dean, D. R.; Huynh, B. H.; Johnson, M. K. *Biochemistry* **2005**, *44*, 12955.

(51) Frees, D.; Qazi, S. N.; Hill, P. J.; Ingmer, H. *Molecular Microbiology* **2003**, *48*, 1565.

(52) Michel, A.; Agerer, F.; Hauck, C. R.; Herrmann, M.; Ullrich, J.; Hacker, J.; Ohlsen, K. *Journal of Bacteriology* **2006**, *188*, 5783.

(53) Tomas, C. A.; Beamish, J.; Papoutsakis, E. T. *Journal of Bacteriology* **2004**, *186*, 2006.

(54) Boschi-Muller, S.; Olry, A.; Antoine, M.; Branlant, G. *Biochimica et Biophysica Acta* **2005**, *1703*, 231.

(55) Sasindran, S. J.; Saikolappan, S.; Dhandayuthapani, S. *Future Microbiology* **2007**, *2*, 619.

(56) Kumsta, C.; Jakob, U. *Biochemistry* **2009**, *48*, 4666.

(57) Poole, L. B. *Archives of Biochemistry and Biophysics* **2005**, *433*, 240.

(58) Guo, K.; Li, L. *Analytical Chemistry* **2010**, *82*, 8789.

(59) Guo, K.; Li, L. *Analytical Chemistry* **2009**, *81*, 3919.

(60) Kang, I. H.; Kim, J. S.; Kim, E. J.; Lee, J. K. *Journal of Microbiology and Biotechnology* **2007**, *17*, 176.

(61) Ha, H. C.; Sirisoma, N. S.; Kuppusamy, P.; Zweier, J. L.; Woster, P. M.; Casero, R. A. *Proceedings of the National Academy of Sciences* **1998**, *95*, 11140.

(62) Brown, A. D. *Biochimica et Biophysica Acta* **1960**, *44*, 178.

(63) Brown, A. D.; Drummond, D. G.; North, R. J. *Biochimica et Biophysica Acta* **1962**, *58*, 514.

(64) Pegg, A. E.; Michael, A. J. *Cellular and Molecular Life Sciences* **2010**, *67*, 113.

(65) Joshi, G. S.; Spontak, J. S.; Klapper, D. G.; Richardson, A. R. *Molecular Microbiology* **2011**, *82*, 9.

(66) Mashima, R.; Nakanishi-Ueda, T.; Yamamoto, Y. *Analytical Biochemistry* **2003**, *313*, 28.

(67) Mailloux, R. J.; Singh, R.; Brewer, G.; Auger, C.; Lemire, J.; Appanna, V. D. *Journal of Bacteriology* **2009**, *191*, 3804.
(68) Mailloux, R. J.; Beriault, R.; Lemire, J.; Singh, R.; Chenier, D. R.; Hamel, R. D.; Appanna, V. D. *Plos One* **2007**, *2*, e690.

Chapter 4. Conclusions and Future work

4.1 Conclusions

My thesis work focused on the application of liquid chromatography mass spectrometry (LC-MS) methods to study the butanol tolerance mechanism of *Staphylococcus warneri* strain SG1 on the proteomic and metabolomic levels. We first profiled the proteome of SG1 grown in the absence and presence of 1-Butanol and compared the proteome expression changes using a semi-quantitative emPAI approach. To achieve more accurate quantification, an isotopic 2-MEGA labeling method was employed that resulted in the identification of 260 differentially expressed proteins of SG1. Detailed information of protein regulation related to butanol tolerance of SG1 was obtained and a metabolomic analysis of amine-, phenol- or carboxylic acid containing metabolites in the TCA cycle was conducted to validate the proteomic results.

In Chapter 2, we described the comparative proteomic profiling of SG1 grown in the presence and absence of butanol using a 2D-LC-MS/MS shotgun approach. A total of 1567 proteins were identified, covering 64 % of the predicted open reading frames. We then applied the ^{15}N isotopic labeling approach to experimentally validate and increase the confidence of spectra assignment, and have constructed an MS/MS spectra library of tryptic digest for SG1 with 3209 unique peptide sequences. Semi-quantitative information on the proteome changes of SG1, based on the emPAI values extracted from MASCOT server, indicated the up-regulation of proteins involved in cell envelop biosynthesis, energy production and oxidative stress response. This is the first comprehensive proteomic profiling analysis of SG1 and provides the basis for understanding the physiology and butanol tolerance mechanism of SG1.

In Chapter 3, we used the 2-MEGA isotopic labeling in combination with 2D-LC-MS/MS to accurately quantify protein expression change of SG1 grown in the presence of butanol. Forward and reverse labeling experiments were performed on biological triplicate samples and differentially expressed proteins were identified with statistical confidence. Pathway analyses indicate that the up-regulated proteins are involved in energy metabolism, oxidative stress response, lipid and cell envelope biogenesis, or have chaperone functions. We then applied differential isotope labeling LC-MS to probe metabolite changes in key metabolic pathways upon butanol stress and found corroborative evidence that the cells were experiencing oxidative stress in the presence of butanol and that changes in the TCA cycle play a key role in the butanol response. This work demonstrates the possibility to combine proteomic and metabolomic analysis together for a more insightful elucidation of butanol tolerance mechanism of SG1.

4.2 Future work

Despite our current understanding of SG1 on both the proteomic and metabolomic levels, there is still plenty of work that needs to be done in order to achieve a more insightful understanding on the butanol tolerance mechanism of SG1. I will briefly describe several aspects as follows.

First, although we extracted hundreds of peak pairs from dansylation labeling and DmPA labeling experiments for the SG1 metabolomics study, metabolite identification still remains to be a challenge considering the lack of authentic standards. In the future, more authentic standards should be used to validate the putative match and to increase the confidence of metabolites identification.

Second, in order to effectively combine proteomic and metabolomic analyses for studying and comparing biological systems, it is vital to increase the metabolites identification coverage. Up to now, we have investigated the amine-, phenol-, and carboxylic acid- containing metabolites but still fail to cover several important metabolic pathways such as glycolysis, pentose biosynthesis pathway

and gluconate interconversion. The metabolites in these three pathways mainly consist of sugars and their derivatives. In the future we will need to apply the differential $^{15}N_2$-/$^{14}N_2$-isotope dansylhydrazine labeling method developed in our group to detect the carbonyl-containing metabolites. With such labeling technique, better metabolite identification coverage, especially on the central metabolic pathways could be obtained and a more comprehensive metabolic map could be established with qualitative and quantitative information on both proteins and metabolites.

Third, although 2-MEGA isotopic labeling combined with 2D-LC-MS/MS provides good coverage and accuracy of differentially expressed proteins of SG1 under different growth conditions, the quantification results are relative and various factors such as instrumental performance, random sampling effect would still affect the overall reliability of the experimental data at least to some extent. Therefore, further validation of the proteomic data using molecular biology approaches would be important for increasing the confidence of the results and eliminating false positives.

Western blot is a standard way for proteomic data validation[1,2], but is limited by the availability of antibodies that recognize the proteins of interest. Alternatively, we can use DNA microarray and real time polymerase chain reaction (RT-PCR) to detect the gene expression, as protein expression is generally correlated to the corresponding transcription strength. DNA microarray is a powerful tool for characterizing gene expression on a genome scale[3], and it has been widely used in discovery-based medical and basic biological research. However, in order to obtain reliable results for validating proteomic data of SG1, a few key issues, including the reproducibility, reliability, compatibility and standardization of microarray analysis and results, must be critically addressed. On the other hand, RT-PCR offers advantages in detection sensitivity, sequence specificity, large dynamic range as well as its high precision and reproducible quantitation compared to DNA microarray and other techniques[4-6]. Extra attention

to the specificity of primer designs is required in order to minimize interference and false positives when applying it in the complex matrix of SG1 samples.

The technologies discussed above have their own strengths and weaknesses for validating proteomic data of bacterial strains. Perhaps, the ultimate way to validate proteomic result is to confirm the biological roles of targeted proteins contributing to butanol tolerance via genetic engineering. This approach involves the overexpression or knockout of one or few genes[7] as well as genomic and global approaches including whole genome shuffling, creating transcription and sigma-factor libraries and deletion libraries, etc[8]. For the purpose of validating proteomic data, overexpression or knockout of critical proteins in the targeted pathway will be more practical approaches. Careful analysis and selection of proteins involved in the important metabolic pathway for butanol tolerance of SG1 would be the key to successfully identify protein candidates that are critical for the butanol-tolerant phenotype of SG1.

In summary, despite the fact that there are still lots of space to improve, the LC-MS approach was successfully applied in my work to get a better understanding for the butanol tolerance mechanism of SG1 on both proteomic and metabolomic level. Consistent findings in metabolomic and proteomic analysis were used to support each other and a comprehensive and detailed understanding in system biology view was achieved. This work also demonstrates the potential and applicability of combining metabolomic and proteomic approaches for the study of various types of biological systems.

4.3 Literature cited

(1) Bantscheff, M.; Lemeer, S.; Savitski, M. M.; Kuster, B. *Analytical and Bioanalytical Chemistry* **2012**, *404*, 939.

(2) Mann, M. *Journal of Proteome Research* **2008**, *7*, 3065.

(3) Wang, Y. L.; Barbacioru, C.; Hyland, F.; Xiao, W. M.; Hunkapiller, K. L.; Blake, J.; Chan, F.; Gonzalez, C.; Zhang, L.; Samaha, R. R. *BMC Genomics* **2006**, *7*.

(4) Arya, M.; Shergill, I. S.; Williamson, M.; Gommersall, L.; Arya, N.; Patel, H. R. H. *Expert Review of Molecular Diagnostics* **2005**, *5*, 209.

(5) Wong, M. L.; Medrano, J. F. *Biotechniques* **2005**, *39*, 75.

(6) Wilhelm, J.; Pingoud, A. *Chembiochem* **2003**, *4*, 1120.

(7) Nicolaou, S. a.; Gaida, S. M.; Papoutsakis, E. T. *Metabolic Engineering* **2010**, *12*, 307.

(8) Warner, J. R.; Patnaik, R.; Gill, R. T. *Current Opinion in Microbiology* **2009**, *12*, 223.