Proteomic analyses of Sulfolobus solfataricus : an extremophilic archaeon
by Richard Cornelius Barry Jr

A dissertation submitted in partial fulfilment of the requirement for the degree of Doctor of Philosophy in Chemistry
Montana State University

Abstract:
Microorganisms which inhabit extreme environments are known as extremophiles. Sulfolobus is a ubiquitous archaeon that can be found in hyperthermophilic and acidophilic habitats, such as the hot springs of Yellowstone National Park.

The proteins an organism uses to cope with environmental variables are known as its functional proteome and are a dynamic physiological response limited by the available genes in the organism. In this dissertation, the techniques of proteomics were applied toward studying Sulfolobus solfataricus.

Using high resolution two-dimensional gel electrophoresis, fluorescent protein stains, gel image analysis software, mass spectrometry, and bioinformatics, the practices and procedures of proteomics were developed. A quantitative and statistical evaluation was used to determine the reproducibility and confidence levels of using gel-separated proteins to monitor cellular protein expression levels. This evaluation strongly indicated the need for statistical rigor when using 2-D gels to determine protein expression differences.

The first proteomic mapping of a Crenarchaeota was performed on Sulfolobus solfataricus P2. A total of 867 protein spots (325 different gene products) from 2-D gels were mapped using MALDI MS and bioinformatic software. The proteomic techniques were extended to Sulfolobus solfataricus P1, where the genome is not known, and it was demonstrated that P1 proteins could be identified using the P2 genomic database, by mapping 420 P1 spots (224 gene products). Some of the proteins identified provide evidence of biochemical pathways that have not yet been explored for this microorganism. Additionally, protein expression differences obtained from using different carbon sources during cultivation revealed the possible presence of a novel respiratory pathway, previously unknown in archaea. Finally, lists of protein candidates have been created which can be used to guide further research to better understand the mechanisms used by Sulfolobus to cope with different carbon sources, or arsenate and arsenite stress.

A foundation has now been created for using a fully integrated proteomics approach to study a wide variety of complex biological systems. The new discoveries about Sulfolobus have already stimulated additional research projects to illuminate the mechanisms involved in its response to environmental variables and can be further used to compliment genomic studies being performed by other research groups.

PROTEOMIC ANALYSES OF *SULFOLOBUS SOLFATARICUS*, AN EXTREMOPHILIC ARCHAEON

by

Richard Cornelius Barry Jr.

A dissertation submitted in partial fulfilment
of the requirement for the degree

of

Doctor of Philosophy

in

Chemistry

Montana State University
Bozeman, Montana

June 2003

APPROVAL

of a dissertation submitted by

Richard Cornelius Barry

This dissertation has been read by each member of the dissertation committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the College of Graduate Studies.

Dr. Edward A. Dratz Edward A. Dratz 6-23-03
(Signature) Date

Approval for the Department of Chemistry and Biochemistry

Dr. Paul Grieco Paul Grieco 6-16-03
(Signature) Date

Approved for the College of Graduate Studies

Dr. Bruce McLeod Bruce R. McLeod 6-30-03
(Signature) Date

STATEMENT OF PERMISSION TO USE

In presenting this dissertation in partial fulfillment of the requirements for a doctoral degree at Montana State University, I agree that the library shall make it available to borrowers under rules of the library. I further agree that copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for extensive copying or reproduction of this dissertation should be referred to Bell & Howell Information and Learning, 300 North Zeeb Road, Ann Arbor, Michigan 48106, to whom I have granted "the exclusive right to reproduce and distribute my dissertation in and from microform along with the nonexclusive right to reproduce and distribute my abstract in any format in whole or in part."

Signature Richard C. Barry

Date 6-23-03

PREFACE

I would like to take this opportunity to thank those individuals who have had a major impact on my success here at Montana State University. This work could not have been accomplished without the support of Drs. Edward Dratz, Mark Young and Tim McDermott. I would like to thank Dr. Edward A. Dratz for being my advisor. It is with great appreciation that I thank Drs. Mark Young and Tim McDermott, Co-Directors of MSU's Thermal Biology Institute, for their financial support of my research projects and educational experiences, and for establishing our successful research collaborations. I would also like to thank my wonderful, fantastic and beautiful wife, Dianne, and beautiful, outstanding, and brilliant daughter, Amanda, for their patience, understanding, support, sacrifice, and love during this long journey.

TABLE OF CONTENTS

TABLE OF CONTENTS, CONT.

Chapter Page

TABLE OF CONTENTS, CONT.

Chapter Page

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES, CONT.

LIST OF FIGURES, CONT.

ABSTRACT

Microorganisms which inhabit extreme environments are known as extremophiles. *Sulfolobus* is a ubiquitous archaeon that can be found in hyperthermophilic and acidophilic habitats, such as the hot springs of Yellowstone National Park.

The proteins an organism uses to cope with environmental variables are known as its functional proteome and are a dynamic physiological response limited by the available genes in the organism. In this dissertation, the techniques of proteomics were applied toward studying *Sulfolobus solfataricus*.

Using high resolution two-dimensional gel electrophoresis, fluorescent protein stains, gel image analysis software, mass spectrometry, and bioinformatics, the practices and procedures of proteomics were developed. A quantitative and statistical evaluation was used to determine the reproducibility and confidence levels of using gel-separated proteins to monitor cellular protein expression levels. This evaluation strongly indicated the need for statistical rigor when using 2-D gels to determine protein expression differences.

The first proteomic mapping of a Crenarchaeota was performed on *Sulfolobus solfataricus* P2. A total of 867 protein spots (325 different gene products) from 2-D gels were mapped using MALDI MS and bioinformatic software. The proteomic techniques were extended to *Sulfolobus solfataricus* P1, where the genome is not known, and it was demonstrated that P1 proteins could be identified using the P2 genomic database, by mapping 420 P1 spots (224 gene products). Some of the proteins identified provide evidence of biochemical pathways that have not yet been explored for this microorganism. Additionally, protein expression differences obtained from using different carbon sources during cultivation revealed the possible presence of a novel respiratory pathway, previously unknown in archaea. Finally, lists of protein candidates have been created which can be used to guide further research to better understand the mechanisms used by *Sulfolobus* to cope with different carbon sources, or arsenate and arsenite stress.

A foundation has now been created for using a fully integrated proteomics approach to study a wide variety of complex biological systems. The new discoveries about *Sulfolobus* have already stimulated additional research projects to illuminate the mechanisms involved in its response to environmental variables and can be further used to compliment genomic studies being performed by other research groups.

CHAPTER 1

PROTEOMICS OVERVIEW

Introduction

The understanding of biological systems at the molecular level has been accelerated in recent years. Until relatively recently, a reductionist approach was followed where each biomolecule of the puzzle was independently studied and painstakingly connected to their interaction(s) with other biomolecules. Using this approach, scientists realized that biological events typically occur through complex networks involving large numbers of molecular interactions. With the advent of new technology, complex biological systems can now be studied from a nonreductionist viewpoint, or global perspective, whereby a great many variables are monitored in parallel. For example, one can simultaneously study the interplay of large numbers of macromolecules under the influence of particular stimulus in order to decipher the players regulated in response to that stimulus. By using a global approach, two specific research areas, Genomics and Proteomics, are increasingly adding to our knowledge about the mechanisms of and evolutionary relationships between many organisms.

The tools and techniques of proteomics are developed and applied in this dissertation to studying the archaeal microorganism, *Sulfolobus solfataricus*. In general, proteomic analysis is more challenging than genomic analysis due to the wide range of properties that proteins have in cells. Robust and reliable methods were established and a

large scale protein mapping was first developed to establish the feasibility of applying proteomics methodology to this organism and to identify the predominate proteins expressed. This was followed by several studies which monitored and characterized the protein expression patterns of *S. solfataricus*, in two different strains and in response to arsenic stress and viral infection. It is hoped that by establishing the proteomic techniques used in this research, other investigators will consider using proteomics as a tool in their research. Further, it is hoped that the findings presented in this thesis will stimulate new research projects which will greatly advance the beginnings made here.

Each type of organism on Earth contains a unique sequence of deoxyribonucleic acids (DNA) that are known collectively as the genome of that organism. Genomic sequences can be interpreted to indicate the gene products that potentially can be expressed by the organism, including the various types of ribonucleic acids (RNA) and a vast number of different proteins. Genomic studies can tell us a lot about the evolutionary lineage of an organism, the proteins it can potentially express, the metabolic pathways it may use, and divulge the characteristics of its DNA replication, transcription, and translation machinery. But not every gene within a genome is expressed nor is the function of every gene predictable; genes that are expressed, are regulated to various levels, depending on environmental and physiological pressures. Further, DNA sequences do not provide information on the rates at which their products are produced or turned over, the functional levels to which they are expressed, or the presence of post-translational modifications which could effect biomolecular interactions and function. Gene expression is a dynamic event that is constantly influenced by the conditions that

surround the organism and, therefore, gene expression is continually being adjusted to balance those conditions. Thus, when studying the physiological responses of organisms to different stimuli, more information can be obtained by monitoring the actual gene products being expressed in response to those stimuli.

mRNA expression analysis is an extremely powerful technique that is yielding impressive new information (Laub et al, 2000). Tools are available which can rapidly identify specific mRNAs based on their unique nucleic acid sequences, as assembled from the four bases, A, C, U, and G. However, tracking mRNA levels has limitations for understanding the mechanisms of biological processes. Recent detailed studies have shown that mRNA levels do not correlate well to protein abundance (Anderson et al, 1997; Gygi et al, 1999; Hochstrasser, 1998), presumably because different mRNAs exhibit different rates of translation and turnover. Gygi and Abersold found that mRNA levels in the liver correlated to their corresponding protein level with a low overall correlation of 0.48 (Gygi et al, 1999). Anderson found that for those mRNAs which had similar expression levels, their translated proteins varied in abundance on average by 30-fold; and for proteins which had comparable abundance, their corresponding mRNA's varied by as much as 20-fold (Anderson et al, 1997). Furthermore, the amount of functional protein (as translated from the mRNA) may be essentially independent of mRNA levels altogether, as a result of post-translational modification(s) (such as phosphorylation or N–acylation) or differential rates of turnover. On the other hand, measuring the levels of proteins directly (along with their isoforms) shows the actual end-

products of genomic responses and provides information on the macromolecules that are most intimately involved with carrying out cellular functions.

Proteomics Defined

The term proteomics was first used in 1995 to describe the protein complement of an organism (James, 2000; Wasinger et al, 1995). Proteomics has now evolved into a completely new scientific discipline and currently cuts across many scientific research areas. Proteomic research is generally focused on one or more of three themes: 1) to elicit the greatest amount of information on the levels of protein expression and changes in protein expression in response to biological stimuli; 2) identify the relationships between proteins, along with their post-translational modifications and cellular response mechanisms; and 3) to characterize the structure and function of proteins (Figure 1). Within all three themes there is a major push to develop high-throughput methods to automate the numerous steps typically required to obtain the data. This includes the need for software and algorithms which can extract the desired information from the voluminous amounts of data that is usually obtained. Upgraded and newer technologies, such as mass spectrometry, X-ray crystallography, and bioinformatics, are being integrated with previously well-established biochemical techniques in order to advance our understanding of how members of the proteome produce the symphony of responses as directed by the genome. A comprehensive description of the proteome of an organism would ideally provide a catalogue of all the proteins expressed by the genome, data on the

changes in protein expression under various defined conditions, the presence of post-translational modifications, and the interactions and distribution of proteins within the cell (Quadroni et al, 1999). The members of the protein orchestra that an organism uses to cope with different stimuli, and their characteristics, are called the functional proteomes for those stimuli. Proteomics has the potential to provide insight into nearly every biological event which occurs within the living cell.

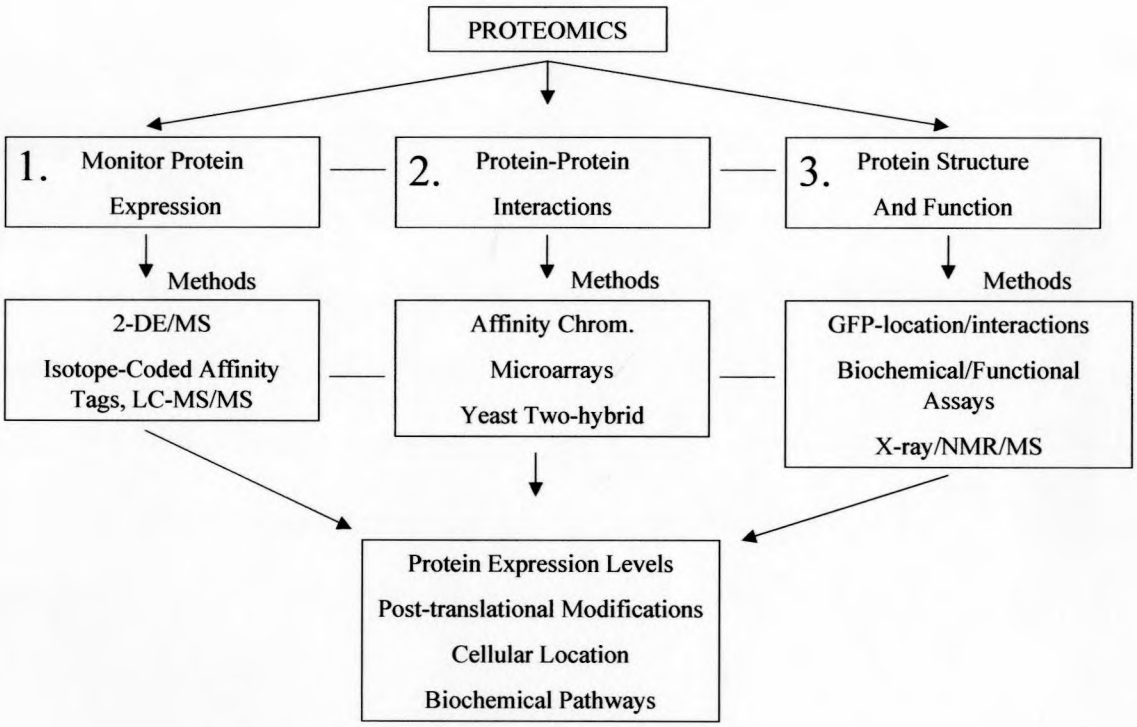

Figure 1. Proteomic approaches. The typical types of proteomic investigations and the types of methods used for analyses: 1) Tracking protein expression levels between two different cellular states or conditions; 2) Determining the interactions of proteins; and 3) Characterizing protein structure. All themes have a common goal of understanding protein functions and roles they play in biochemical pathways. (Two-dimensional gel electrophoresis, 2-DE; Mass spectrometry, MS; Liquid Chromatography and multidimensional MS, LC-MS/MS; Nuclear Magnetic Resonance, NMR; X-ray crystallography, X-ray)

Detailed Objectives of this Research Project

Montana State University recently established the Thermal Biology Institute (TBI) (www.tbi.montana.edu) to study the rich thermophilic communities in Yellowstone National Park (YNP). The technique of proteomics was developed and applied in this thesis to study an organism known to inhabit the thermal hot springs around the world, including YNP. Specifically, the hyperthermophilic and acidophilic archaeon, *Sulfolobus solfataricus* was used as a model system because it naturally inhabits many of the hot springs of YNP, can be cultured in a laboratory setting to high cell densities, has had its genome fully sequenced, and because a proteomic approach had never before been used to study this unique organism. The aim of this project was to develop the techniques of proteomics and show how this method can be used to make new scientific discoveries about extremophilic microorganisms. My general goal was to develop robust methods that were highly reproducible and reliable, so that they may be used as a foundation for future investigations. The effectiveness of proteomics was illustrated by determining the protein expression profiles of this organism under various conditions, as briefly summarized in the objectives below.

Objective 1: Develop facile and reliable two-dimensional gel (2-D gel) separation methods for the proteins in *Sulfolobus solfataricus*. Protein expression patterns were determined for *Sulfolobus solfataricus* strain P2 as cultured in the laboratory. Typically, it has been difficult to resolve all of the proteins in cells on 2-D gels, but this resolution is improved by subfractionation of the proteins into two or more classes prior to isoelectric

focusing. The approach maximizes resolution and provides information on the subcellular location of the expressed proteins. The fractions used were: 1) aqueous tris-soluble proteins and 2) aqueous tris-insoluble proteins. Aqueous tris-soluble extracts should contain mostly cytosolic proteins, but may also include some peripheral membrane proteins. Aqueous tris-insoluble extracts should contain integral membrane and peripheral membrane proteins. After separating and resolving the classes of proteins on 2-D gels, many proteins were excised, proteolytically digested, and analyzed by matrix assisted laser desorption/ionization time of flight mass spectrometry (MALDI MS). It was critical to confirm our ability to accurately and definitively identify proteins from 2-D gels using the mass-to-charge ratios (molecular weight) obtained from peptides of protein digests and the software available for matching experimental peptides to theoretical peptides (as deduced from the genome of *Sulfolobus solfataricus* P2). Next, methods were developed to extract proteins from whole cells of *Sulfolobus*, without fractionation into classes, to facilitate later experiments (see objectives 3 and 4). The most effective 2-D gel methods were established using analytical evaluations, and based on gel-to-gel reproducibility, resolution, the numbers of proteins detected, and the robustness of the method.

Objective 2: Determine if the same protocols and analyses can be applied to *Sulfolobus solfataricus* P1, a different strain of *S. solfataricus*, and perform strain comparisons at the level of the proteome between the two organisms. The study of strain P1 was judged to be important because it is generally more susceptible to certain types of viral infections (personal communication, Ken Stedman). Currently, the genomes of *S.*

solfataricus P2 and *S. tokodaii* have been sequenced, while the genome of P1 has not. It is unknown how similar P1 is to P2. It will be determined if the P2 protein database can be used to identify P1 proteins. By viewing the actual proteins expressed in each strain, we may be able to reveal differences that future investigators can use to design studies which explore more fully the meaning of those differences.

Objective 3: Perform comparative proteomic analyses on *S. solfataricus* P2 when exposed to different concentrations of arsenite (As III, 350 μM and 1mM) and arsenate (As V, 350 μM and 1 mM). Some of the solfataric hot springs, such as those in YNP, are known to contain high concentrations of arsenic which are toxic to most organisms. The ability to cope with harsh environmental variables, such as arsenic, presumably requires the existence of unique cellular machinery. Examining the proteomic strategies and mechanisms used by *Sulfolobus* to survive in the presence of such toxic metaloids has implications in evolution and may spotlight proteins worth further investigation. Difference gel electrophoresis was employed to reveal those proteins significantly up- or down-regulated in response to the metaloids. Highlighted proteins were excised and identified using MALDI MS and database search algorithms. A list of the regulated proteins is presented which identifies those proteins that the regulatory system of the organism adjusts in the presence of arsenic stress. It is anticipated that these results will stimulate additional research centered on the toxic metal resistance system(s) of this organism.

Objective 4: Investigate *S. solfataricus* P1 response to the *Sulfolobus shibatae* virus, SSV2, as a pilot study for more in-depth studies in the future. *Sulfolobus* has been

found to host a variety of viruses and plasmids. These extrachromosomal elements have been determined to be unique in DNA sequence. Currently, it is unknown exactly how *Sulfolobus* is infected, how it responds to infection, the extent to which host proteins and metabolic pathways are affected by infection, and how virion assembly and export occur. Using the proteomic methods developed in Objective 3, the approach will be used to investigate host response to viral infection and the results used to develop research directions to further understand those results. The results of this objective will be used as an example of another way to apply proteomics toward studying *Sulfolobus,* has not been completed at the time of this writing, and will be attached as an appendix.

Future Prospects to Build on the Foundation of this Research

The research in this dissertation will hopefully lay the foundation for future proteomic research projects on extremophiles and microorganisms worldwide. All of the tools used in this study, from two-dimensional gel electrophoresis, gel image analysis, automated MALDI mass spectrometry data acquisition, to database searches have been well-tested to insure their successful application by future proteomic investigators. As an attempt to help the success of future projects, a mass spectrometry user's guide (for automatic MALDI MS data acquisition and data processing, User Guide for AutoXecute) and a two-dimensional gel image analysis user's guide (using PDQuest from Bio-Rad, PDQuest, A User's Guide for Beginners) have been created and are currently available through Prof. Edward A. Dratz, Department of Chemistry and Biochemistry. Given the