



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

© Copyright by Richard Cornelius Barry Jr (2003)

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

D378
B2796

APPROVAL

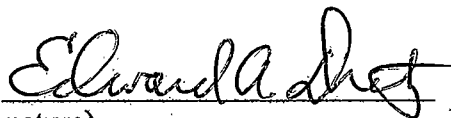
of a dissertation submitted by

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

(Signature)



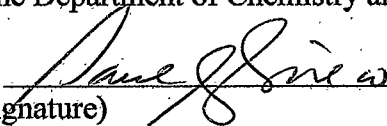
Date

6-23-03

Approval for the Department of Chemistry and Biochemistry

Dr. Paul Grieco

(Signature)



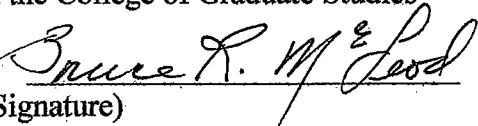
Date

6-16-03

Approved for the College of Graduate Studies

Dr. Bruce McLeod

(Signature)



Date

6-30-03

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

Chapter	Page
1. PROTEOMIC OVERVIEW	1
Introduction	1
Proteomics Defined	4
Detailed Goals and Objectives of the Research Project	6
Future Prospects to Build on the Foundation of this Research	9
The New "Omic" Sciences	10
2. EXTREMOPHILES, ARCHAEA, AND SULFOLOBUS	13
Extremophiles	13
The Three Domains of Life	14
Archaea	14
Archaeal Lipids	16
The S-Layer and Pseudomurein	19
Archaeal Introns	23
Background on <i>Sulfolobus</i>	24
Codon Usage	30
Proteomic Studies can Compliment Genomic Findings	32
3. PROTEOMIC TOOLS	34
Classical Proteomic Approach	35
Two-Dimensional Gel Electrophoresis, 2-DE: Carrier Ampholytes	35
Immobilized pH Gradient IEF	39
New Reagents for 2-DE	40
Equilibration for Second-Dimension Separations	43
SDS-PAGE	44
Visualizing Proteins on 2-D Gels	46
Image Analysis Software	48
The Need for Mass Spectrometry	49
MALDI TOF MS	50
Non 2-DE Proteomic Approaches	51
Ion-Exchange Chromatography	51
Reversed Phase High Performance Liquid Chromatography	53
Size-Exclusion Chromatography	56
Electrospray Mass Spectrometry	57
Peptide Mass Fingerprinting	62

TABLE OF CONTENTS, CONT.

Chapter	Page
4. QUANTITATIVE EVALUATION OF SAMPLE APPLICATION METHODS FOR SEMI-PREPARATIVE SEPARATIONS OF BASIC PROTEINS BY 2-DE	67
Chapter Summary	67
Introduction	68
Electrophoresis Equipment and Chemicals	72
Sample Preparation	73
Cell Culture and Harvesting	73
Extraction of Proteins and Preparation for IEF	74
Isoelectric Focusing and Sample Application	75
Cup-Loading	76
Active Rehydration	78
Passive Rehydration	78
Equilibration and SDS-PAGE	79
Gel Imaging	80
Replicate Gels and Image Analysis	80
Spot Detection and Quantitation	81
Results	82
Comparison of the Number of Spots Detected	82
Correlation Coefficients, Spot Matching, and Matching Efficiency	87
Affects of Sample Preparations and Repeat Gels on Matching	92
Reproducibility of Spot Quantity within a Sample Loading Method	93
Spot Quantity Comparisons between Sample Loading Techniques	96
Discussion	99
Concluding Remarks	103
Addendum	104
5. PROTEOMIC MAPPING OF THE EXTREMOPHILIC ARCHAEON SULFOLOBUS SOLFATARICUS P2	109
Chapter Summary	109
Introduction	110
Electrophoresis Equipment and Chemicals	116
Cell Culture and Harvesting	116
Differential Solubilization of Proteins and Preparation for IEF	116
Extraction of Tris-Soluble Proteins	116
Extraction of Tris-Insoluble Proteins	117
Two-Dimensional Gel Electrophoresis	118
Spot Detection and Image Analysis	119

TABLE OF CONTENTS, CONT.

Chapter	Page
In-Gel Trypsin Proteolysis and Peptide Extraction	120
Peptide Mass Fingerprinting	121
Results and Discussion	122
Concluding Remarks	139
6. COMPARISONS BETWEEN SULFOLOBUS SOLFATARICUS	151
STRAINS P1 AND P2 USING 2-DE AND MASS SPECTROMETRY	
Chapter Summary	151
Introduction	152
Materials and Methods	153
Cell Culture and Harvesting	153
2-DE, Spot Detection, Image Analysis and Database Searches	154
Results: Proteomic Mapping of P1 Proteins	154
Results: Functional Proteome Comparisons between P1 and P2	167
Discussion	174
Concluding Remarks	181
7. PROTEOMIC RESPONSE OF SULFOLOBUS SOLFATARICUS P2	184
TO ARSENITE AND ARSENATE	
Introduction	184
Electrophoresis Materials and Equipment	188
Method Development	188
Results and Discussion	189
Arsenite Exposure	189
Arsenate Exposure	189
Spot Harvesting and Spot Identification	190
Concluding Remarks	190
8. SUMMARY OF RESEARCH ACCOMPLISHMENTS	196
APPENDIX A - PROTEOMIC CHARACTERIZATION OF VIRAL	202
INFECTION IN ARCHAEA: APPLICATION TO	
<i>S. SOLFATARICUS</i> P1 AND SSV2	
LITERATURE CITED	210

LIST OF TABLES

Table	Page
1. Extremophiles	13
2. A comparison of traits between archaea, bacteria, and eucarya	16
3. Sensitivity of gel staining methods	47
4. Number of spots detected using different sample loading methods	85
5. Correlation coefficients between replicate gels for each sample loading method ...	88
6. Average coefficient of variation for landmarked spot quantities	95
7. Comparison of protein quantities in landmarked spots	97
8. Quantity values for landmarked spots from sample A	105
9. Quantity values for landmarked spots from sample B	106
10. Quantity values for landmarked spots from sample C	107
11. Quantity values for landmarked spots from sample D	108
12. P2 tris-soluble and tris-insoluble proteomic mapping	141
13. P1 tris-soluble and tris-insoluble proteomic mapping	161
14. P1 vs. P2 difference gel analysis results, proteins uniquely detected in P1/P2 ...	171
15. Proteins uniquely detected and their biochemical pathways	172
16. Blastp sequence homology results of homologous bacterial plasmids	187
17. Identification of proteins regulated on IPG 4-7 gels to arsenic	193
18. Identification of proteins regulated on IPG 6-11 gels to arsenic	195

LIST OF FIGURES

Figure	Page
1. Proteomic approaches	5
2. Proteomics is the protein complement expressed by an organism	10
3. The “omic” sciences	12
4. The universal tree of life	15
5. Archaeal lipids	18
6. The structure of peptidoglycan	21
7. The structure of pseudopeptidoglycan	22
8. Electron microscope image of a cross-section of <i>S. solfataricus</i> P2	26
9. Protein length has evolutionary and protein folding implications	29
10. Amino acid composition in <i>S. solfataricus</i> P2	30
11. The classical proteomic approach	36
12. Two-Dimensional gel electrophoresis	38
13. Immobilization of ampholytes	39
14. Types and properties of new reagents for 2-DE	42
15. Structure and properties of detergents	43
16. Percentage of acrylamide and molecular weight separation ranges	45
17. Polymerization of acrylamide and bis-acrylamide	45
18. Example of a multiply charged series for a protein	59
19. Deconvolution of the multiply charged series	59
20. A homemade assembly for cup-loading on the Protean IEF Cell	76

LIST OF FIGURES, CONT.

Figure	Page
21. Representative 2-D gel images	84
22. Cross-section image of <i>S. solfataricus</i> P2	112
23. Labeled image of P2 tris-soluble IPG NL 3-10, 8-18%, in urea and CHAPS	124
24. Labeled image of P2 tris-soluble IPG NL 3-10, 8-18%, in urea, thiourea, CHAPS, and ASB-14	125
25. Labeled image of P2 tris-soluble IPG 4-7, 8-18%, in urea and CHAPS	127
26. Labeled image of P2 tris-soluble IPG 6-11, 8-18%, in urea and CHAPS	128
27. Labeled image of P2 tris-insoluble IPG NL 3-10, 8-18%, in urea, thiourea, CHAPS, and ASB-14	130
28. Labeled image of P2 tris-insoluble IPG 4-7, 8-18%, in urea, thiourea, CHAPS, and ASB-14	131
29. Labeled image of P2 tris-insoluble IPG 6-11, 8-18%, in urea, thiourea, CHAPS, and ASB-14	132
30. MALDI MS of PEPCK	137
31. Database search results for PEPCK	138
32. Labeled image of P1 tris-soluble IPG NL 3-10, 8-18%, in urea and CHAPS	155
33. Labeled image of P1 tris-insoluble IPG NL 3-10, 8-18%, in urea, thiourea, CHAPS, and ASB-14	156
34. Labeled image of P1 tris-insoluble IPG 4-7, 8-18%, in urea, thiourea, CHAPS, and ASB-14	157

LIST OF FIGURES, CONT.

Figure	Page
35. Labeled image of P1 tris-insoluble IPG 6-11, 8-18%, in urea, thiourea, CHAPS, and ASB-14	158
36. Comparison of tris-soluble and tris-insoluble IPG NL 3-10 P1 with P2	169
37. Comparison of tris-soluble IPG 4-7 P1 with P2 and comparison of tris-soluble IPG 6-11 P1 with P2	170
38. Induction of CODH and N-methylhydantoinase	173
39. PDQuest image analysis of 350 μ M arsenite and 350 μ M arsenate exposure	191
40. PDQuest image analysis of 1mM arsenite and 1 mM arsenate exposure	192
41. The viruses of <i>Sulfolobus</i>	208
42. P1 response to SSV2	209

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.

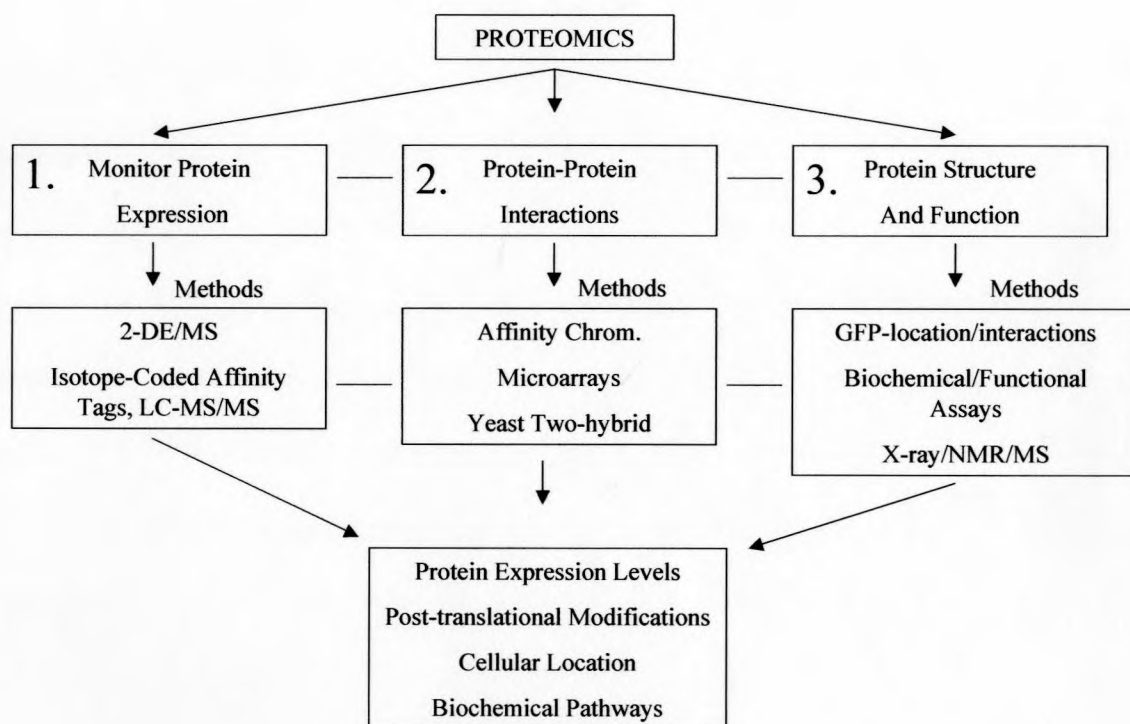


Figure 1. Proteomic approaches. The typical types of proteomic investigations and the types of methods used for analyses: 1) Tracking protein expression levels between two different cellular states or conditions; 2) Determining the interactions of proteins; and 3) Characterizing protein structure. All 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.*

solfatarius 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 μ M and 1mM) and arsenate (As V, 350 μ M 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

campus-wide level of interest in the technologies used in this project, it is anticipated that the work presented in this thesis will provide a foundation for new avenues of discovery in many research areas.

The New “Omic” Sciences

Proteomics is one path to discovery in the biosciences and, like many fields of research, can be used to complement other types of studies. Most importantly, proteomics can complement genomic studies by providing information on the actual gene products expressed, the extent of expression, the relationships between expression products and stimuli, the presence of protein post-translational modifications, and can guide investigations on the structure and function of gene products, to name a few (Figure 2).

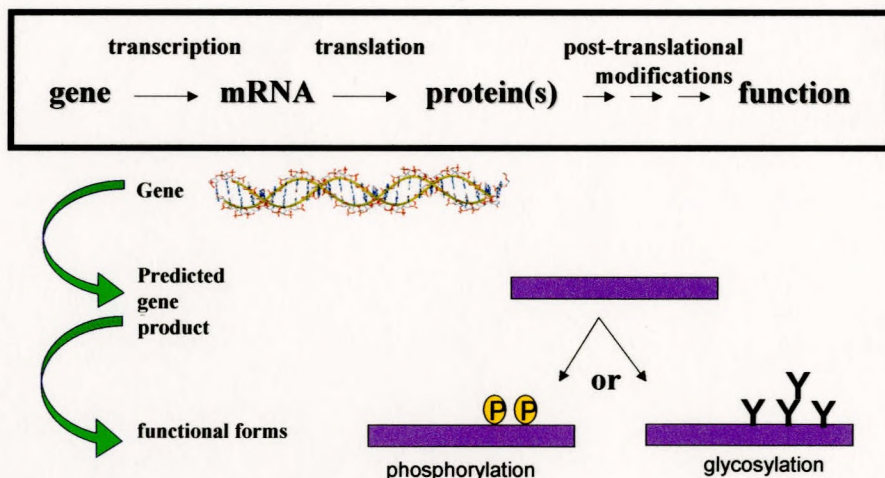


Figure 2. Proteomics is the protein complement expressed by an organism. Proteomics complements genomic studies by characterizing the actual end-products of genomic responses to stimuli and the presence of post-translation protein modifications required to carry out cellular processes.

These areas of research have become so popular that various forms of genomics and proteomics are popping up everywhere, with each having its own “-omics” catch phrase. Figure 3 shows the relationship between some of the more popular “omic” sciences. Genomics is focused on sequencing the DNA of an organism, annotating the genes, mapping the location of the genes in chromosomes, and predicting the types of proteins expressed, along with their functions. The ultimate goal of genomic studies is to better understand the relationship between genes, their normal function, and role in cellular processes. Functional Genomics helps achieve this goal by monitoring gene activity under different cellular states; that is, which genes are turned on or off and determining the function of the gene. The functionality or lack of functionality of a gene can provide information about the possible relationship(s) to other genes. Gene activity and function are usually investigated by using mutagenesis or by monitoring the types of mRNA and proteins being expressed.

Correlating integrated networks of gene activity to phenotypic behavior, however, is a much more complicated task. For example, synergistic interactions between gene products and alterations in the post-translational modifications (PTMs) of proteins can often produce dramatic physiological responses even though there is relatively little or no detectable change in gene activity. Thus, the complementary technique of proteomics is often used to illuminate the protein players, their PTMs, and corresponding genes required to produce the response. Proteomics can be subdivided into three broad areas of research: Structural Proteomics, Comprehensive Proteomics, and Functional Proteomics. Structural Proteomics includes studies which compare the three-dimensional structures of

proteins across different species. Structural Proteomics can yield information on the function and evolutionary history of gene products. Comprehensive Proteomics typically uses high-throughput methods to analyze as many proteins as possible from one organism, with the goal of further characterizing the properties of an organism. Functional Proteomics is used to monitor the expression levels of proteins, including those post-translationally modified, to study cellular response to stimuli. In addition to the forms listed above, there are new types of “omic” sciences which are continually christened to describe the study at hand. Metabolomics, for example, investigates the changes observed in small molecule metabolites. Thus, with all the new “omics” popping up, there was temptation to title this dissertation “Sulfolomics”.

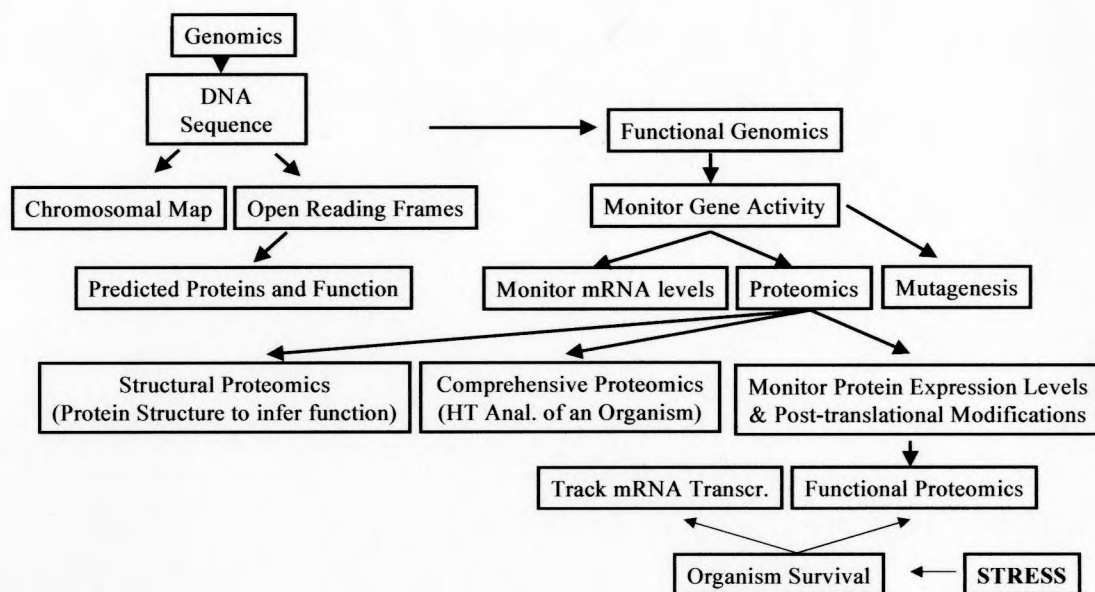


Figure 3. The “omic” sciences. Genomics and Proteomics have various subgroups and the relationship between some of the “-omics” are represented above.

CHAPTER 2

EXTREMOPHILES, ARCHAEA, AND SULFOLOBUS

Extremophiles

It is amazing how life finds ways to exist under incredible extremes of temperature, pH, pressure, dessication, and salinity. Recent discoveries of microorganisms capable of thriving in boiling hot-springs, volcanic vents, acid mine drainages, and within Antarctic rocks have motivated efforts to understand the corresponding biological strategies and mechanisms used in order to support life under such conditions. These organisms are known as extremophiles and are classified according to the nature of their habitats, as shown in Table 1.

Table 1. Extremophiles (Hough et al, 1999)

<u>Phenotype</u>	<u>Environment</u>	<u>Genera</u>
Thermophilic	55-80°C	<i>Methanobacterium</i> , <i>Thermoplasma</i> , <i>Thermus</i> , <i>Bacillus</i>
Hyperthermophilic	80-113°C	<i>Aquifex</i> , <i>Archaeoglobus</i> , <i>Hydrogenobacter</i> , <i>Methanothermus</i> , <i>Pyrococcus</i> , <i>Pyrodictium</i> , <i>Pyrolobus</i> , <i>Sulfolobus</i> , <i>Thermococcus</i> , <i>Thermoproteus</i> , <i>Thermotoga</i>
Psychrophilic	-2 to 20°C	<i>Alteromonas</i> , <i>Psychrobacter</i>
Halophilic	2-5 M NaCl	<i>Haloarcula</i> , <i>Halobacterium</i> , <i>Haloferax</i> , <i>Halorubrum</i>
Acidophilic	pH < 4	<i>Acidianus</i> , <i>Desulfurolobus</i> , <i>Sulfolobus</i> , <i>Thiobacillus</i>
Alkaliphilic	pH > 9	<i>Natronobacterium</i> , <i>Natronococcus</i> , some <i>Bacillus</i> species

Since the structures of ribosomal RNA molecules tend to be conserved in certain regions, it is possible to align the base sequences of the relatively homologous regions in order to determine which positions vary and the number of changes that can be distinguished between any two organisms (White, 2000). This allows phylogenetic and evolutionary relationships to be deduced as a result of the differences that exist.

The Three Domains Of Life

The current phylogeny of life indicates that there are three domains: Archaea, Eubacteria (bacteria), and Eucarya (eucaryotes) (Figure 4). 16S Ribosomal-DNA sequencing and other studies have established that many of these extremophiles represent a unique domain of life known as Archaea (formerly Archaeobacteria), which are distinct from Bacteria (eubacteria) and Eucarya (eucaryotes) (White, 2000). While many extensive studies have been undertaken on extremophiles over the last 25 years, the designation of Archaea as a unique domain of life has propelled the recent excitement in extremophilic research (DeRosa, 1996).

Archaea

Archaea and bacteria are both considered procaryotes and share many characteristics. They vary in size from 0.5-2 μM in diameter and have a relatively large outer surface area relative to internal cell volume (a ratio much larger than found in

eucaryotic cells). This large surface-to-volume ratio allows the exterior of the microbe to maximize its ability to absorb and transport nutrients into the cell while minimizing the time to distribute them within the small volume. Prokaryotes do not contain organelles (such as a membrane bound nucleus, mitochondria, or chloroplasts) but metabolic functions are still compartmentalized and take place in multienzyme granules, in plasma membranes, or within the periplasm of gram-negative bacteria (White, 2000).

Prokaryotes can also contain special compartments known as inclusion bodies which carry specific enzymes, storage products, or photosynthetic pigments to perform specialized tasks (White, 2000).

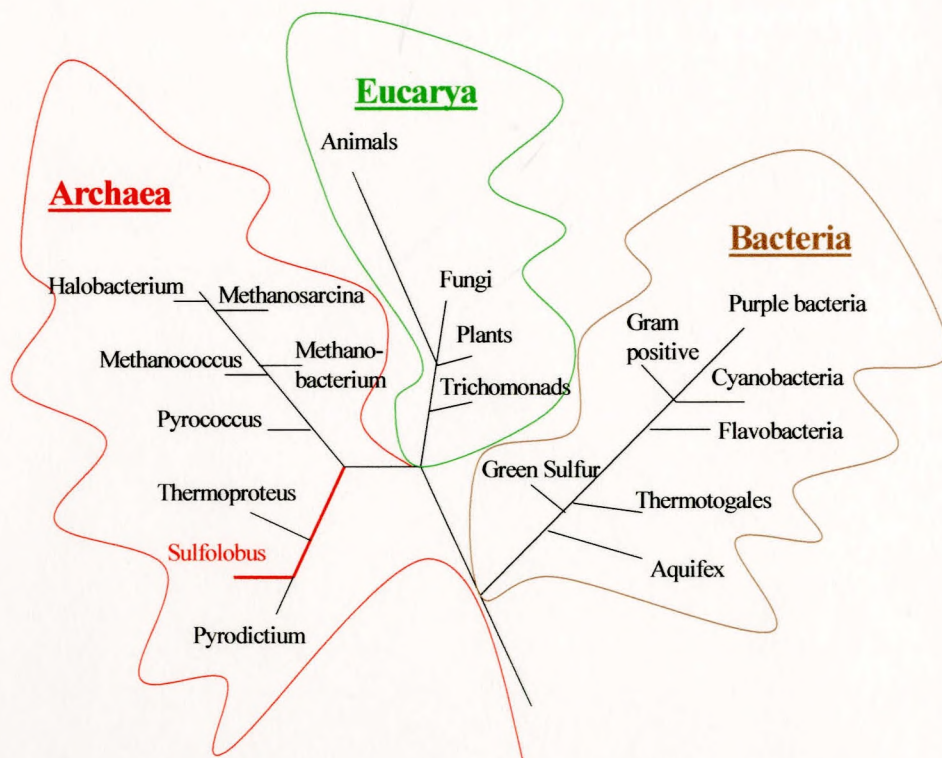


Figure 4. The Universal Tree of Life showing the three domains: Archaea, Bacteria, and Eucarya. (Adapted from van de Vossenberg, *et al.*, 1998). Eucarya and bacteria are subdivided into divisions, while archaea are separated according to genus.

Archaea and bacteria may also contain pili appendages for cell adhesion and flagella for swimming. Currently, there is no evidence that procaryotes contain cytoskeletal networks to help them maintain cell shape, but there have been suggestions that cytoskeleton-like structures do exist (Trent et al, 1994; White, 2000). The general structure of archaeal and bacterial DNA are also similar in that the DNA molecule (or chromosome) is a circular closed loop (although new findings are modifying this connection). In contrast, eucaryotic cells have numerous organelles, each carrying out specialized functions, and linear DNA molecules (Table 2).

Table 2. A comparison of traits between Archaea, Bacteria, and Eucarya		Eucaryotic Cells				
Structure	Function	Archaea	Bacteria	Fungi	Plants	Animals
Nucleus	Membrane-bound, contains DNA	-	-	+	+	+
Mitochondria	ATP energy, respiratory	-	-	+	+	+
ER *	Processing and transport of proteins	-	-	+	+	+
Ribosomes	Translation of DNA for protein synthesis	+	+	+	+	+
Cilia	Cell movement	+	+	-	-	+
Flagella	Cell movement	+	+	+	+	+
Cytoplasmic Membrane	Semipermeable membrane	+	+	+	+	+
CW + Peptidoglycan*	Cell strength and rigidity, protection against osmotic shock	-	+	-	-	-
CW - Peptidoglycan*	Protects against osmotic shock, physical strength	+	-	+	+	-
Pseudomurein	Cell strength and rigidity	+	-	-	-	-
Introns	Noncoding regions of the genome, transcribed, then excised	+	-	+	+	+
* ER: endoplasmic reticulum, CW: cell wall						

Archaeal Lipids

Archaea differ dramatically from the other two domains of life in cell membrane composition, cell wall chemistry, cellular processes, and in 16S ribosomal RNA sequences. Archaeal membrane lipids, such as those of the hyperthermophile, *Sulfolobus solfataricus*, consist of isoprenoid alcohols which are ether-linked to glycerol(s) to form

di- and tetraethers (Figure 5) (DeRosa, 1996; van de Vossenberg et al, 1998). Bacterial and eucaryotic glycerides, on the other hand, are fatty-acid esters linked to glycerol. The tetraether lipids of archaea form cross-links between the headgroups and create very rigid membranes that are nearly impermeable to protons and ions. *Sulfolobus*, for example, contains two glycerol headgroups linked together by two biphytanyl moieties (Hopmans et al, 2000; Luzzati et al, 1987). More importantly, the ether cross-linked lipids cannot be easily degraded, are mechanically and temperature resistant, and tolerant to osmotic shock (Albers et al, 2001; van de Vossenberg et al, 1998; van de Vossenberg et al, 1999). Salt tolerance and low ion permeability may be a result of the organization of the lipid (Albers et al, 2001; van de Vossenberg et al, 1995; van de Vossenberg et al, 1999). Electrical conductance of the archaeal membrane was determined to be very low indicating that these membranes are good insulators and can account for their low ion permeability (Gulik et al, 1988; Hanford et al, 2002). An organizational model of the tetraether lipid membrane is still emerging but it is likely that the hydrophobic core is responsible for the low ion permeability. Some extremophiles, like *Halobacterium*, contain primarily diethers, while most of the sulfur-dependent thermophiles have primarily tetraethers and a trace of diethers (White, 2000). Other archaea have a mixture of ethers.

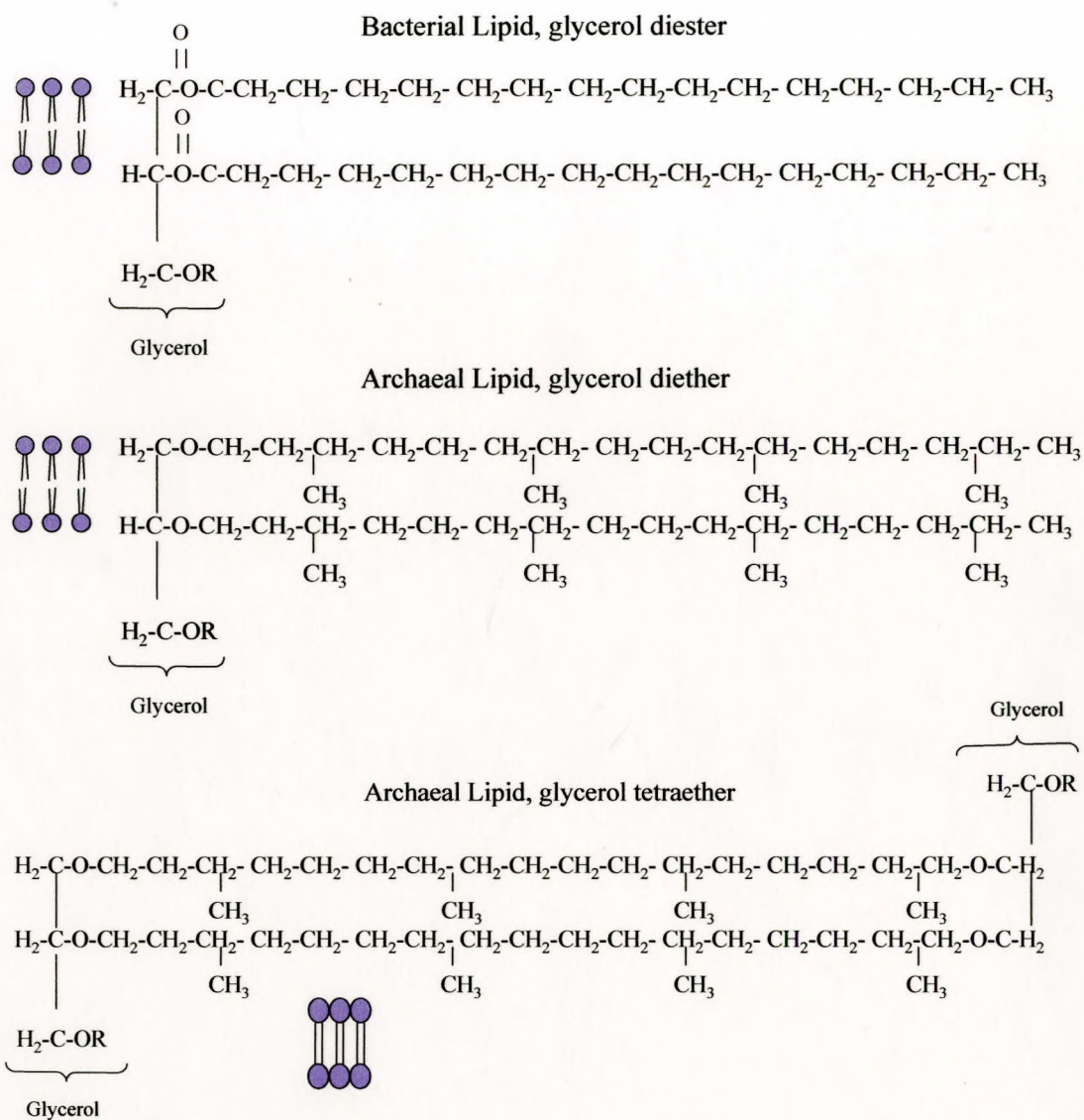


Figure 5. Archaeal lipids. The lipids that make up archaeal cytoplasmic membranes consist of glycerol di- and tetraethers. Bacterial and eucaryotic cytoplasmic membranes are made of glycerol diesters. R: represents polar groups, such as phosphate and choline. Purple dots represent polar R-groups attached to the lipids and indicated the structure of the bilayer formed.

The S-layer and Pseudomurein

Many extremophilic archaeons also possess a crystalline, self-assembled surface layer of proteins, known as the S-layer (Sara et al, 2000; Schuster et al, 2000). S-layers are not unique to archaea or always found in extremophiles, but when present these layers seem to impart special properties which protect those that have them from environmental influences. In bacteria, S-layers have been found to function as adhesion sites for exoenzymes (Egelseer et al, 1996; Mayr et al, 1996), deter predations by *Bdellovibrio* bacteriovirus (Koval et al, 1991), and act as specific sites for virion attachment or adsorptions (Sara and Sleytr, 2000). *Sulfolobus* has a firmly attached S-layer of arrayed proteins on its surface and, hence, this layer may also be employed as a protection mechanism.

The chemical composition of archaeal cell walls are unique from other domains of life. Cell walls are generally not homogeneous structures, but are an assembly of layers of different combinations of proteins, polysaccharides, lipids, and other molecules. Two different types of walls exist among the bacteria: gram positive and gram negative, which are described to provide perspective on the archaeal walls. Whether a bacterium is classified as gram positive or negative depends upon the thickness of its peptidoglycan layer. Peptidoglycans are cross-linked peptides and sugars (glycans). More specifically, peptidoglycan is a cell wall polymer consisting of an alternating assembly of N-acetylglucosamine (GlcNAc) and N-acetylmuramic (MurNAc), attached via β 1,4 linkages (Figure 6). Additionally, tetrapeptides are attached to each MurNAc and the

peptides from different MurNAc's are linked together, creating the support structure for the bacterial cell wall. The peptides are usually L-alanine, D-glutamate, a diamino acid (meso-diaminopimelic acid), and D-alanine. Gram positive bacteria have a thick peptidoglycan layer and retain crystal violet-iodine during Gram staining. Gram negative bacteria have a thin layer and consequently lose the crystal violet stain during the organic washing step of the staining procedure. Archaea lack a peptidoglycan layer and, therefore, cannot be technically classified as either gram positive or negative. Instead, archaeal cell walls contain a pseudopeptidoglycan structure (also known as pseudomurein), polysaccharides, and proteins (the S-layer mentioned earlier) (Sara and Sleytr, 2000). The same sugar, GlcNAc, is present in archaeal cell walls but it is cross-linked via $\beta 1,3$ linkages to N-acetylalosaminuronic acid (also known as pseudomurein) instead of N-acetylmuramic acid using $\beta 1,4$ linkages (Figure 7). In addition, the peptides which link the pseudomurein together are all L-amino acids whereas bacterial peptidoglycans contain D-amino acids. Collectively, the affects of an S-layer, pseudoglycan layer, and glycerol di- and tetraether linked lipids give extremophiles like *Sulfolobus* a relatively tough protective shell that it can use to withstand extreme environments.

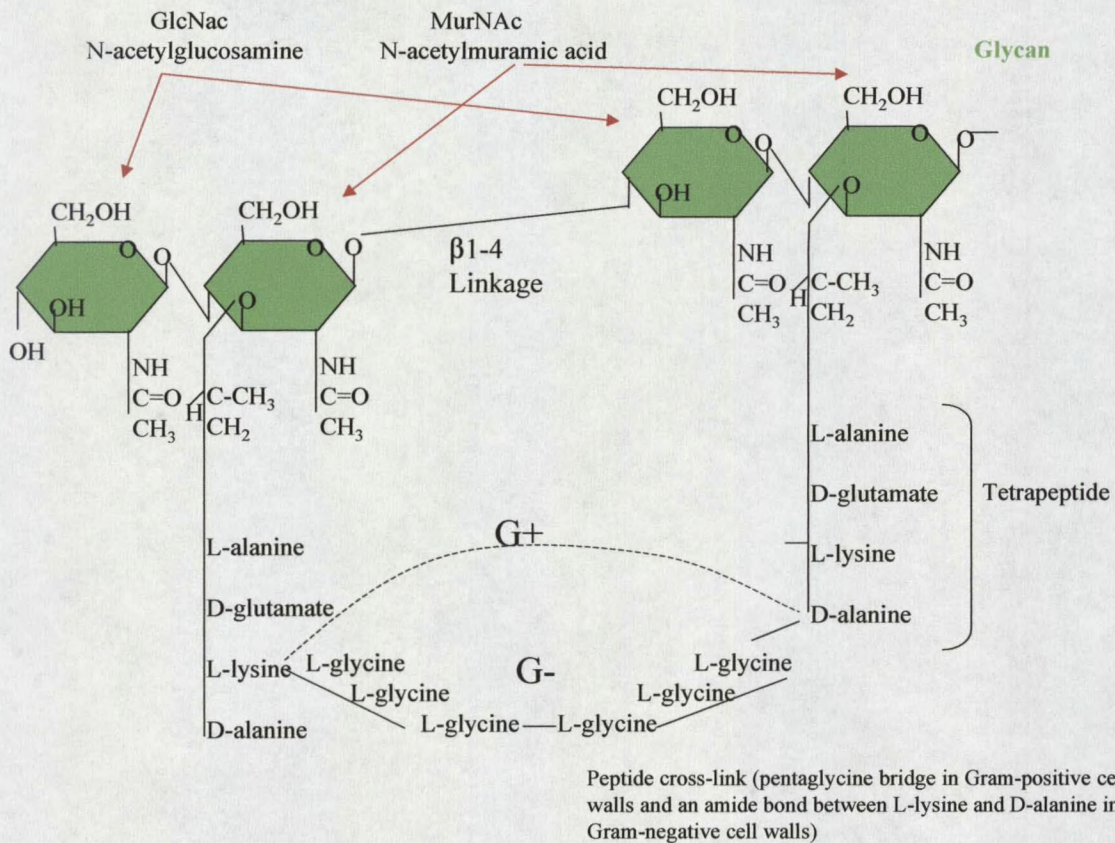


Figure 6. Peptidoglycan backbone of the bacterial cell wall. It is composed of repeating alternating units of N-acetylglucosamine and N-acetylmuramic acid. The tetrapeptides of N-acetylmuramic acid form cross-links to provide the structural support of the cell wall.

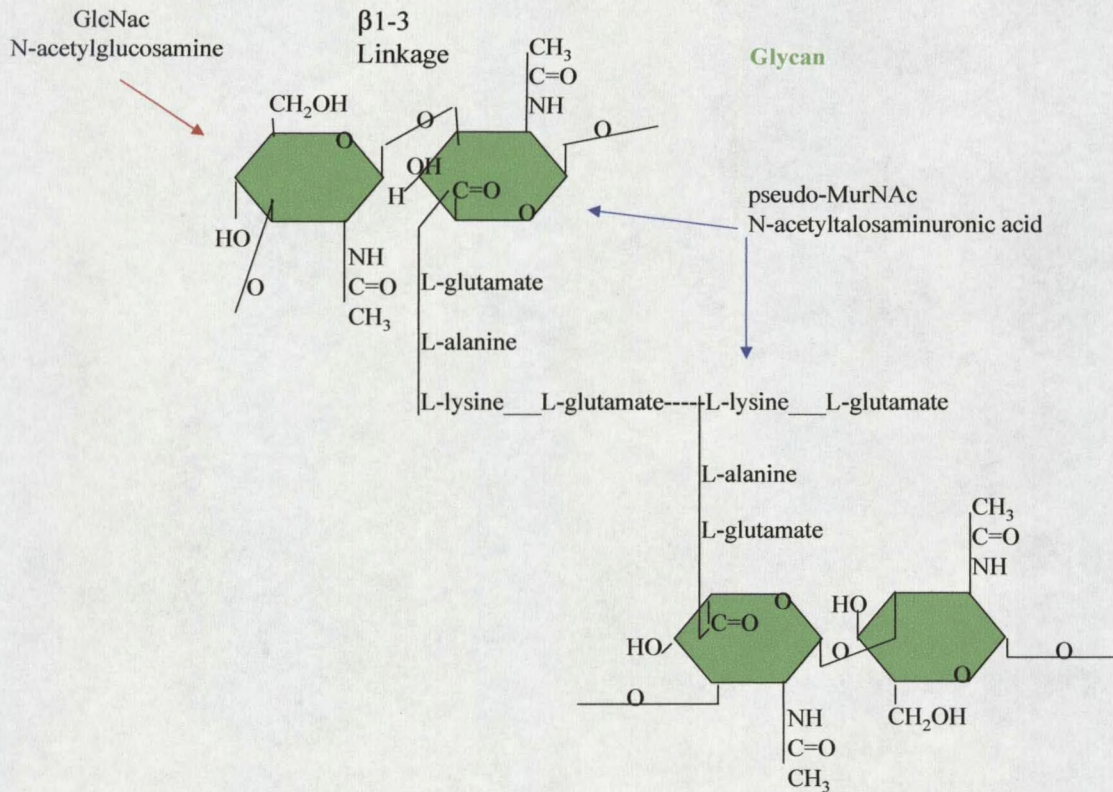


Figure 7. Archaeal cell walls are composed of peptide cross-linked glycans (known as pseudopeptidoglycan) from N-acetylglucosamine and N-acetyltalosaminuronic acid. Only D-amino acids are used in the peptide chains and the glycosidic linkage between GlcNAc and N-acetyltalosaminuronic acid is β 1-3 instead of β 1-4.

Eucaryotic cells are devoid of a peptidoglycan layer like archaea and bacteria. The plasma membrane of the eucaryotic cell is a continuous bilayer of phospholipids containing various embedded proteins. For structural integrity, plant and animal cells use an array of protein filaments which form a complex network within the cytoplasm, known as the cytoskeleton.

Archaeal Introns

Another important characteristic of archaea is the presence of a limited number of introns within the DNA. Introns are noncoding regions which are transcribed into RNA but are then excised by RNA splicing to produce mature RNA. Introns are ubiquitous in eucaryotic genes but are rarely found in bacteria. In 2002, two landmark papers were published which clearly demonstrated the presence of a unique class of introns in the pre-RNA of archaea (*Aeropyrum pernix*, *Archaeoglobus fulgidus*, *Sulfolobus solfataricus* and *Sulfolobus tokodaii*) (Tang et al, 2002; Watanabe et al, 2002). Introns were found either in tRNAs (mostly within the anticodon loop and in a few instances within the anticodon stem and variable loop) or they were found at multiple locations within the large rRNA in the organisms used in the study (Watanabe, et al, 2002).

Additional findings show that archaea are similar to eucaryotes in that they have several kinds of RNA polymerases, a large number of histone-like proteins, DNA packaged in the form of nucleosomes, and that translation is initiated with methionine instead of formyl-methionine (in bacteria) (White, 2000). Thus, archaea share

characteristics which are similar to both bacteria and eucarya. Archaeal DNA replication, transcription, translation, ribosomal proteins, chaperones, and histone-like proteins (known to associate with histones) have eucaryotic traits, while their metabolic processes are more similar to bacteria.

This is truly an exciting time to be studying extremophilic archaeons. Currently, the genomes of 16 archaeons have been completed and many more are in the pipeline (www.tigr.org). Investigations into the mechanisms employed by archaea, in response to extreme environmental variables and other factors (e.g. viral infections), promise to provide a wealth of information as to the cellular constituents required for adaptation, the processes of archaeal gene expression, and perhaps will lead us to reevaluate the nature and evolution of life as we know it

Background on *Sulfolobus*

Phylogenetic analyses of the Archaeal domain of life divide it into two lineages: Euryarchaeota and Crenarchaeota (DeLong et al, 2001; Hershberger et al, 1996; Pace, 2001). A third lineage, Korarchaeota, has also been proposed, but currently consists of only a few sequences and additional data is needed (Barns et al, 1996; Reysenbach et al, 2000). Euryarchaeota are comprised of a heterogeneous group of organisms which include the extreme halophiles and methanogens, the sulfur-dependent hyperthermophile, *Thermococcales*, and the thermophilic/acidophilic Thermoplasmales (Granoff et al, 1999). Crenarchaeota are more phenotypically homogeneous with members being sulfur-

dependent hyperthermophiles and thermophiles (Hershberger, et.al., 1996).

Crenarchaeota include the anaerobic hyperthermophile, *Thermoproteales*, and the hyperthermophilic/acidophilic aerobe, *Sulfolobus*. Currently, Sulfobales represent archaea with the most eucaryotic-type features as suggested by the presence of introns, the structure of their 50S rRNA, mechanisms of their genetic machinery, and the presence of proteins belonging to the family of histone deacetylases (Seeger et al, 1993; Sensen et al, 1998; She et al, 2001).

The genus *Sulfolobus* was first described by Brock in 1972 as a sulfur-oxidizing "bacteria" containing lobed cocci, capable of growing at high temperature and low pH (Figure 8) (Brock et. al., 1972). It is now known that this archaeon can grow much faster on organic substrates than chemolithotrophically and that a variety of strains exist (Brock T.D. et al, 1972; Schafer, 1996). The genus includes a long list of related species, which most notably includes *S. solfataricus*, *S. acidocaldarius*, *S. brierleyi*, *S. shibatae*, *S. metallicus*, and *S. tokodaii*. *Sulfolobus* strains have been isolated from Yellowstone National Park (USA), El Salvador, New Zealand, Italy, Iceland, the West Indies, and Japan. The natural habitats of *Sulfolobus* are usually geothermal areas where elemental sulfur is abundant, especially in solfataric hot springs.

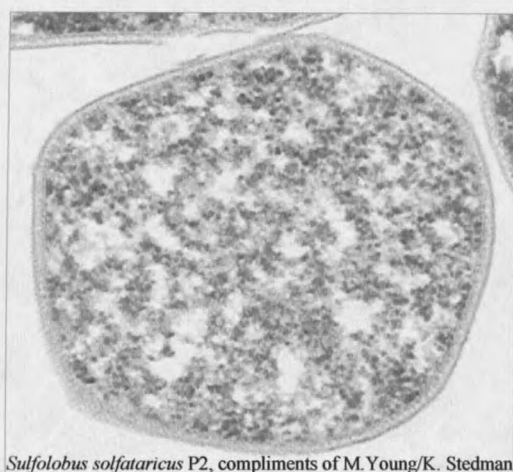


Figure 8. Electron microscope image of a cross-section of *S. solfataricus* P2.

These springs can range in pH from 0.5-6 and have temperatures from 60 °C to nearly 100 °C, with optimum *Sulfolobus* growth occurring at 80 °C and pH 3 (Brock et.al., 1972).

Currently, *S. solfataricus*, *S. acidocaldarius*, and *S. tokodaii* have been the most studied, because of their ability to grow aerobically and heterotrophically to high cell densities under simulated native conditions. Moreover, the complete genome of *S. solfataricus* P2 (2.99 Mbp) and *S. tokodaii* strain 7 (2.69 Mbp) have been published (Kawarabayasi et al, 2001; She et al, 2001). *S. solfataricus* P2 has been sequenced by a joint Canadian-European consortium and the sequence, along with related information is available at: <http://www-archbac.u-psud.fr/projects/sulfolobus/> as well as deposited in GenBank, accession number: AE006641. *S. tokodaii* strain 7 was sequenced by a Japanese group and is available at:

http://b-yahiko.bio.nite.go.jp:8080/dogan/MicroTop?GENOME_ID=st_G1, accession number: AB022438. *In silico* proteomic comparisons between *S. solfataricus* and *S. tokodaii* can be found at EMBL-EBI: <http://www.ebi.ac.uk/proteome/index.html>.

Sulfolobus solfataricus P2 contains 2,992,245 base-pairs on a single circular closed loop (She et al, 2001). The total G+C content is 36%. The open reading frames (ORFs) of the genome encode 2977 proteins, of which 33% have unknown function (hypothetical proteins). ORFs were assigned according to the level of confidence obtained from the algorithm used for comparing *Sulfolobus* proteins with those in the public databases (Nelson et al, 1999). ORFs were assigned only when a Level 1 (90-100% match) or Level 2 (80-89% match) confidence was obtained (Gaasterland et al, 1994). Based on amino acid sequence homology, 40% of the predicted proteins were homologous specifically to archaea, 12% were exclusive to bacterial protein matches, and 2.3% were specific to eucaryotic protein matches. Twenty-three percent (700 ORFs) contained homology to both bacteria and eucarya (She et al, 2001). Fifty-two different gene families were identified, with the largest members containing proteins involved in "fatty-acid" biosynthesis (acetyl-CoA synthetases), alcohol and other dehydrogenases, and ATP transporters. There are 43 tRNA genes with 18 containing introns. Fourteen of the 18 introns were also found in other archaea, suggesting these introns are unique to archaea and not specifically this extremophile (She et al, 2001).

By comparison, *Sulfolobus tokodaii* strain 7 has 2,694,756 bp in its chromosome, with a G+C content of 33% (Kawarabayasi et al, 2001). There are 46 tRNA genes and 24 of them contain introns. ORFs encode for 2,826 proteins with 33% of the them being hypothetical proteins. ORFs were assigned using the Smith-Waterman alignment algorithm (Kawarabayasi et al, 2001). Approximately 91% of the ORFs encode for proteins with 100 amino acids or more, while 268 encode for proteins containing between

50-100 amino acids. Protein size is relevant because it has implications in the evolutionary relationship between archaea, bacteria and eucarya. The average protein length of archaea was found to be very similar to average protein lengths found in bacteria, while archaeal proteins were roughly half the average length of proteins found in eucaryotes. It has been proposed that protein length can be correlated to protein structure and that the "compactability" of a protein can be used to explain why certain structures are common among the domains of life. The average amino acid length for *S. solfataricus* is shown in Figure 9 in comparison to *S. tokodaii*. From a database search using EMBL's European Bioinformatics Institute website (<http://www.ebi.ac.uk>), protein length comparisons were made between archaea, bacteria, and eucarya.

It was once thought that the GC content of extremophiles would necessarily be high in order to produce greater DNA thermal stability through hydrogen bonding (three H-bonds per G or C versus only two H-bonds per A and T). Genomic studies have thus far proven this theory false (De Vendittis et al, 1996; Tekaia et al, 2002). It turns out that the GC content is much lower than expected (Tekaia et al, 2002). The same thought process was also applied to the amino acid content of extremophiles. It was suggested that certain types of amino acids may be required for protein stability in the presence of physical and geochemical extremes. However, no such abundance patterns have been found among the amino acids (Jaenicke et al, 1998; Tekaia et al, 2002). The overall amino acid content of *S. solfataricus* P2 is shown in Figure 10. Instead, it appears that protein stability is affected by noncovalent interactions, such as salt bridges and local hydrophobic bonding (Bohm et al, 1994; Jaenicke and Bohm, 1998; Ladenstein et al, 1998).

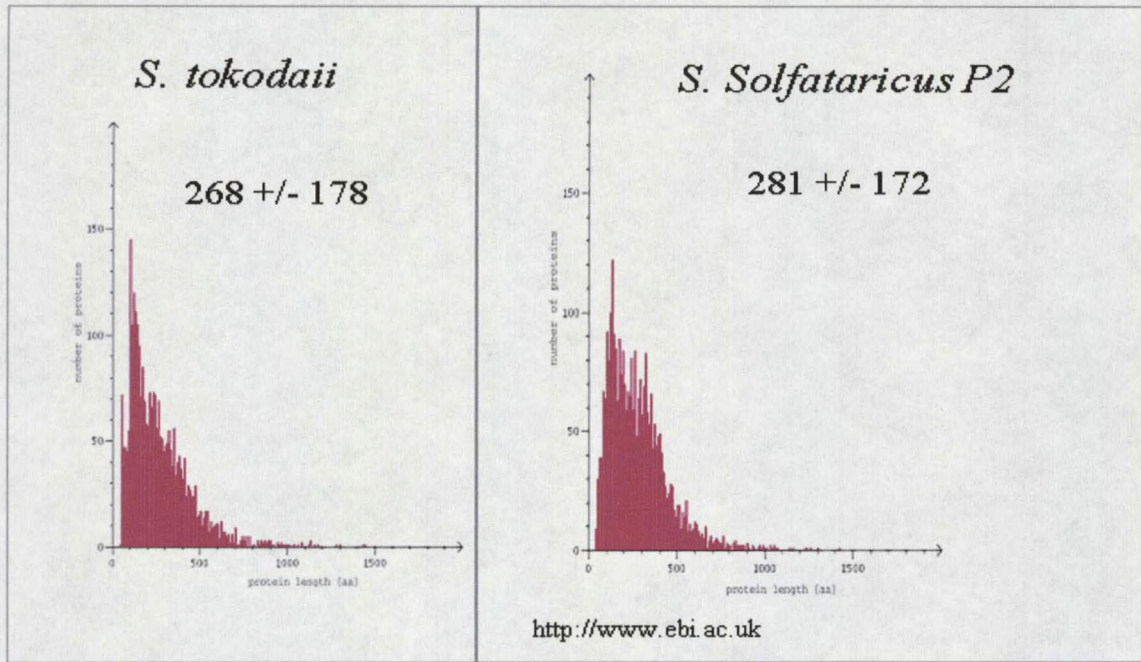


Figure 9. Histograms of the number of proteins (y-axis) of *S. tokodaii* and *S. solfataricus* P2 and the number of amino acids (length, x-axis) they contain. The average protein length for archaea is similar to bacteria and ~ 50% of those found in eucarya. Protein length has evolutionary and protein folding implications.

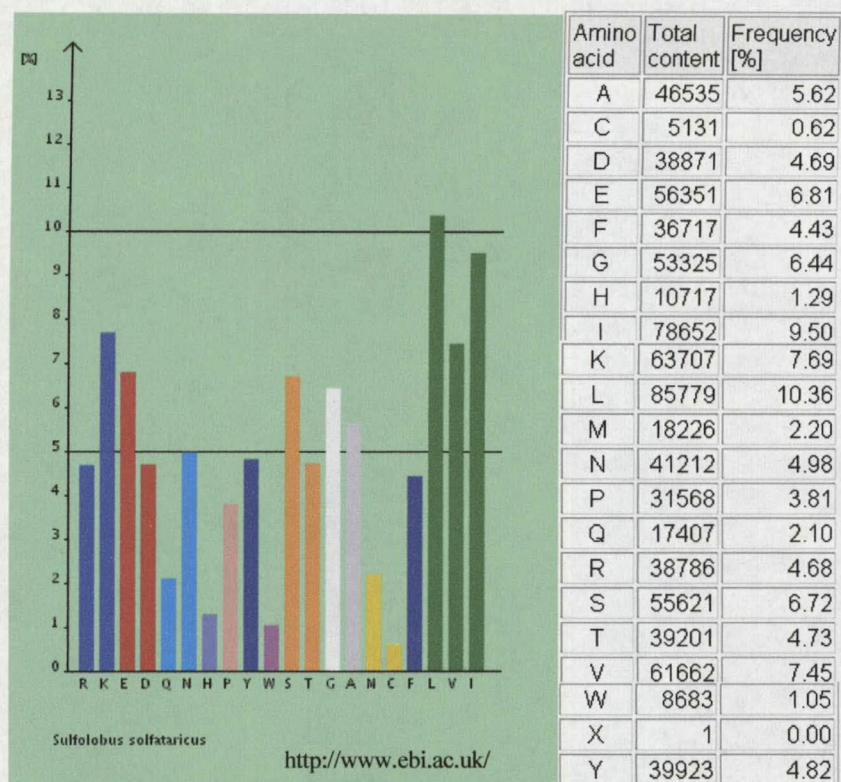


Figure 10. Amino acid composition in *S. solfataricus* P2 based on the ORF's from gene sequencing.

Codon Usage

Codon usage in *Sulfolobus* and comparisons to *Pyrococcus*, another sulfothermophile, was reported by De Vendittis and Bocchini (De Vendittis and Bocchini, 1996). Their study was based on comparisons between 71 total genes from *Sulfolobus acidocaldarius* and *Sulfolobus solfataricus* and 23 total genes from *Pyrococcus furiosus* and *Pyrococcus woesei* (as selected from proteins previously characterized from these organisms). Codon preferences for all four were found to correlate with their low GC

(high AT) compositions. In *Sulfolobus* and *Pyrococcus*, the usage of alternate start codons is significant (De Vendittis and Bocchini, 1996). From a study of 144 genes from *S. solfataricus*, Tolstrup, *et al.*, showed that 79% start with ATG, 11% with GTG, and 10% with TTG (Tolstrup et al, 2000). Furthermore, a start codon was recognized 79% of the time, if it was the first possible start codon following an upstream stop codon, and 19% if it was the second start codon, followed by 3% if it was the third. Amino acid analyses also suggested that *Sulfolobus* and *Pyrococcus* had a three letter gene code propensity to start with A and end in T, which corresponds to their high AT content. These two sulfothermophiles also contained a very low percentage of cysteine residues (De Vendittis and Bocchini, 1996). Using EMBI's website, the overall cysteine content for *S. solfataricus* P2 is 0.62%, for *S. tokodaii* is 0.65%, and for *P. furiosus* is 0.59% (www.ebi.ac.uk). *Sulfolobus* and *Pyrococcus* displayed very similar codon usage, suggesting a close taxonomic relationship. And just as interesting, their codon usage showed numerous similarities to the eucaryote, *Saccharomyces cerevisiae* (yeast), and bovine mitochondria; while clear dissimilarities to the bacterium, *Escherichia coli*, and most surprisingly dissimilar to the halophilic archaeon, *Halobacterium*, were identified.

In addition to analyzing the amino acid content of similar proteins between *Sulfolobus* and *Pyrococcus*, unrelated proteins were evaluated in terms of hydrophobicity (De Vendittis and Bocchini, 1996). In general, all of the unrelated proteins between the two sulfothermophiles shared a proclivity for hydrophobic amino acids. A comparison of the hydrophobic nature of the amino acids in these thermophiles with those of mesophiles (organisms which grow between 20-45°C) and halophiles (extremophilic organisms which

grow in 2-5 M salt) demonstrated that hydrophobic residues were prevalent only in thermophiles (De Vendittis and Bocchini, 1996). This may mean that more hydrophobic amino acids are needed by thermophiles to enhance protein stability.

Sulfolobus is also known to host numerous viruses and plasmids (also known as extrachromosomal elements, see appendix, page 200) (Rice et al, 2001; Zillig et al, 1998). For *Sulfolobus*, these viruses and plasmids are unique in their genetic sequences, not sharing any sequence homology to any other known viruses, archaea, bacteria, or eucarya. Current efforts are under way to take advantage of these extrachromosomal elements by developing genetic shuttle vector systems (Aagaard et al, 1996; Aravalli et al, 1997; Stedman et al, 1999a). These systems would allow scientists to better understand archaeal gene expression, as well as create techniques for expressing proteins for biochemical and structural studies.

Proteomic Studies can Complement Genomic Findings

Scientists can create a tentative list of all of the proteins that an organism can potentially express from its genome. From the genetic information, details about the physiology, metabolic pathways, motility, DNA structure, DNA cell cycle, modes of DNA repair, replication, transcription, and translation can be used to determine the machinery that is available to the organism. Proteomic studies can reveal and monitor the expression of that machinery during the life cycle of the cell. Besides determining which gene products are available, genomics can also tell us which products are absent. This

information can be useful in explaining experimental results and the physiological response of an organism. It also can be used to limit the types of experiments performed to only those which have merit to begin with. By applying proteomic techniques, we should be able to confirm the presence and track the dynamic expression levels of various metabolic pathways, the production of various replication, transcription, and translation machinery, and energy yielding systems. It is believed that, while genomic DNA limits the assortment of gene products available, it is the tracking of the dynamic protein expression patterns that can lead us more directly toward understanding biological response to stimuli. Certainly, the more we study organisms from both a proteomic and genomic perspective, the more we will learn about evolution and the strategies and mechanisms employed by microbes to adapt to environmental stresses.

CHAPTER 3

PROTEOMIC TOOLS

The confluence of bioinformatics and the development of new mass spectrometric instruments have given birth to a whole new way of investigating complex biological systems. The speed and processing power of computer software has been incorporated with mass spectrometric tools that can detect biomolecules in the attomole (10^{-18}) to zeptomole (10^{-21}) range (Belov et al, 2000; Burlingame et al, 1990; Gygi et al, 2000; Loo et al, 1999a; Nguyen et al, 1995; Quadroni et al, 1999; Roepstorff, 2000). That means we now have the capability of detecting and identifying proteins which are present at very low copy numbers per cell. In addition to these tools, an ever-evolving selection of complimentary instruments, such as liquid handling and gel-spot harvesting robots, have made “doing” proteomics less cumbersome and, therefore, are advancing this field. It is understood that well-established techniques became “well-established” because they are well-proven over time. In this research endeavor, new protocols were developed and used to provide important new information on the biological system, *Sulfolobus solfataricus*.

The proteomic approach used in this dissertation falls under the Comprehensive and Functional Proteomic definitions. My goal was to globally monitor (qualitatively and quantitatively) *Sulfolobus* protein expression in order to identify those proteins normally expressed from those which are up- or down-regulated in response to different stresses, and to interpret the mechanism of response as implied from the regulated proteins. In order to accomplish this, one must have the ability to separate and resolve up to ~3000

different cellular proteins, track and compare protein expression, positively identify proteins, and be able to interpret protein regulation in order to identify protein networks where possible.

Classical Proteomic Approach

The well-established technique of two-dimensional gel electrophoresis (2-DE) has been greatly improved over the last 30 years and now has the dynamic range and resolving power to contribute greatly to studying the proteome of this organism. Typical 2-D gel formats of 20 x 20 cm can resolve between 1000 and 3000 proteins. Larger formats on the order of 40 x 30 cm have been used to resolve up to 10,000 proteins (James, 2000). A resolution of 0.01 pH units in pI and 1000 Da in molecular mass is routinely achieved, and conditions can be adjusted to do better if needed. Classical proteomic tools include isoelectric focusing (IEF), polyacrylamide gel electrophoresis (PAGE), matrix-assisted laser desorption ionization time of flight mass spectrometry (MALDI TOF MS), and bioinformatics. Classical proteomics can also include reversed phase high performance liquid chromatography (RP-HPLC) and electrospray ionization mass spectrometry (ESI MS/MS). A typical classical proteomics approach is shown in Figure 11.

Two-dimensional Gel Electrophoresis, 2-DE: Carrier Ampholytes

Two dimensional gel electrophoresis (2-DE) separates proteins on the basis of

charge and size. It takes advantage of the fact that proteins are amphoteric molecules with characteristic masses and charges. A mixture of proteins is initially separated according to their isoelectric point in the first dimension which is known as isoelectric focusing (IEF). IEF is accomplished by maintaining a pH gradient with ampholytes or by covalently attaching buffering groups in a gel matrix. IEF is then coupled with polyacrylamide gel electrophoresis in the presence of the anionic detergent sodium dodecyl sulphate (SDS-PAGE) and proteins are further separated in a second dimension according to size (Karger et al, 1996) (Figure 12).

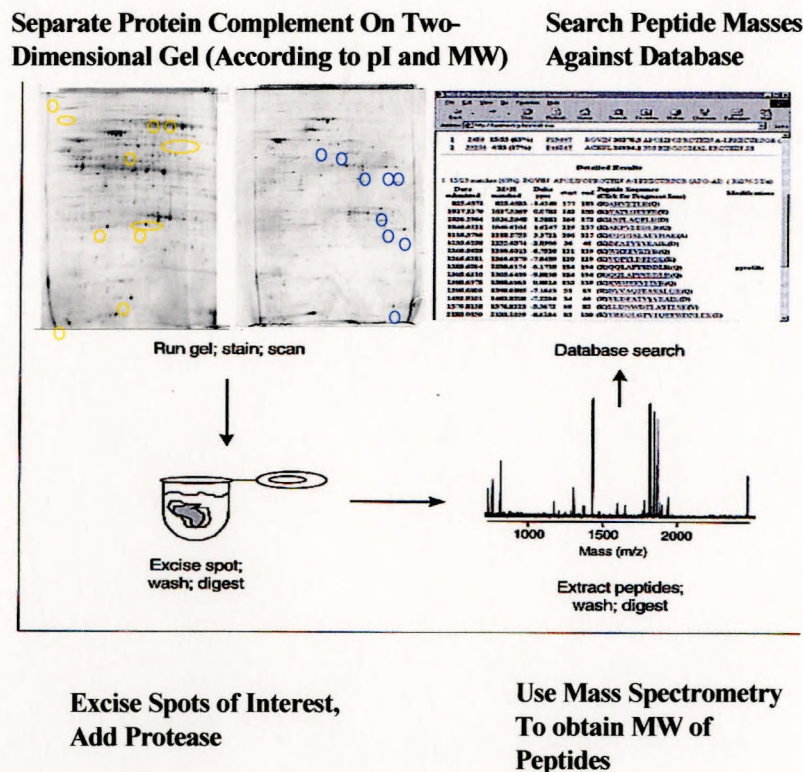


Figure 11. The classical proteomic approach uses two-dimensional gel electrophoresis, image analysis software, MALDI TOF MS, and database searching algorithms to separate cellular proteins, distinguish those proteins that are up- or down-regulated in response to stimuli and to definitively identify the proteins affected.

Since its initial inception by O'Farrell (1975), well over a 1000 papers have been published using 2-DE. But it has only been in the last 12 years that improvements in the tools and techniques used for 2-DE have made it a robust and more reproducible method. Original isoelectric focusing was performed in thin polyacrylamide tubes (130 x 2.5 mm i.d.) with a porosity of ~ 4-8%. Before any protein extract was added, a pH gradient was created using carrier ampholytes (CA). CA are charged molecules consisting of numerous amines (positively charged groups) and carboxylic acids (negatively charged groups). The sum total of amines and carboxylic acids gives each CA its characteristic pI (or net charge). When a mixture of carrier ampholytes, having a range of pI's, are placed within an electric field they migrate towards either the anode (if the net charge is negative) or the cathode (if positively charged). As they migrate they either lose protons or gain protons until they establish a net charge of 0 and reach the "steady-state". If a carrier ampholyte or protein should diffuse from its pI, it will immediately gain charge and be electrophoretically refocused. In the process of proton gain or loss, a pH gradient is established within the solution, which is used to resolve one protein from another. This entire process takes place within polyacrylamide tubes or strips. The use of carrier ampholytes in this fashion is known as CA 2-DE. Unfortunately, the use of CA in IEF is subject to "cathodic drift", whereby pH gradients previously established by prefocusing with ampholytes slowly change with time, causing irreproducible gels. This technique has been relegated to specialized separations (e.g. the nonequilibrium separation of extremely basic proteins like histones and ribosomal proteins). However, all of the IEF principles are applicable to the new technique of immobilized pH gradient focusing.

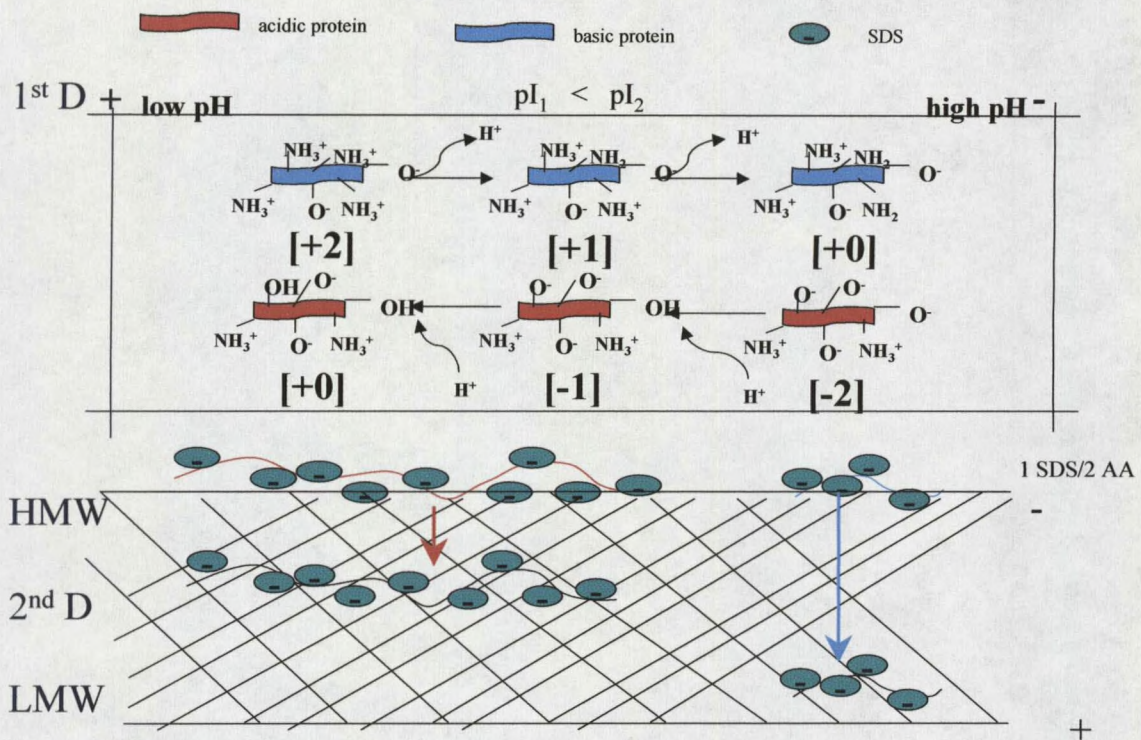


Figure 12. Two-dimensional gel electrophoresis. First dimension separation of proteins is according to charge and takes place within specially prepared polyacrylamide strips. The charge on a protein is represented by its unique isoelectric point, pI , and is determined from its amino acid sequence. In an electric field, acidic proteins migrate toward the anode (+) and absorb protons from the solution until the net charge on the protein becomes 0; and basic proteins migrate toward the cathode (-), giving off protons, until they have a net charge of 0. Once focused, the proteins are linearized and negatively charged by coating with SDS, known as equilibration. The equilibrated proteins are then drawn out of the specialized focusing strips and separated through a molecular sieve made of porous polyacrylamide, completing a two-dimensional separation. See text for more detail.

Immobilized pH Gradient IEF

The IEF method of choice involves immobilizing carrier ampholytes by covalently attaching them to an acrylamide matrix (Figure 13). This is referred to as immobilized pH gradient isoelectric focusing, IPG IEF (Bjellqvist et al, 1982; Gorg et al, 1988) and is a much more reproducible technique. IPG gels are now usually poured onto stiff backings instead of using tubes which has made first-dimension gels much more mechanically strong and easier to use. Since CA are fixed in strips, gel-to-gel reproducibility (as affected by first dimension separations) has dramatically improved. The creation of IPG strips is probably the single most important factor as to the resurgence of 2-DE. Moreover, different combinations of carrier ampholytes can be used to create a selection IPG strips with well defined pH ranges over predetermined distances. This includes linear pH ranges or non-linear pH ranges, broad pH ranges (seven or more pH units), short-ranges (three pH units), and narrow pH ranges (one pH unit) created for large and small format gel use.

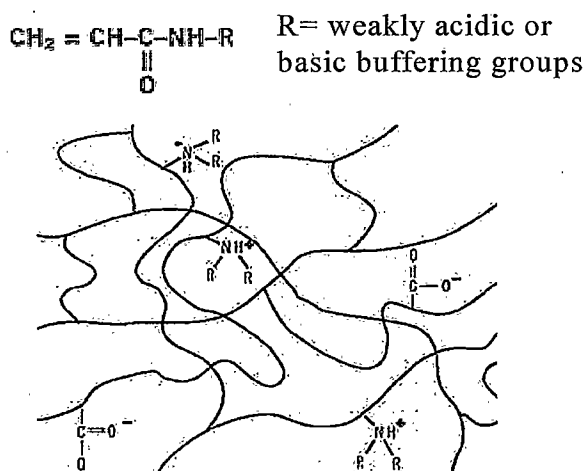


Figure 13. Immobilized carrier ampholytes create regions within the polyacrylamide with a characteristic pKa and buffering capacity (Gorg et al, 1988).

One might find, for example, that many proteins within a mixture have pI's that are too similar to be resolved using a linear pH gradient. Thus, one could either use a non-linear pH gradient strip which contains a very shallow pH gradient in one part of the strip or use a narrower pH range strip consisting of a narrow range of immobilized carrier ampholytes (covering the same distance of gel, end-to-end) in order to resolve the proteins with very similar pIs in the mix.

IPG strips are now commercially available from a variety of vendors and in a variety of lengths and pH gradients. IPG strips are poured in batches at the companies and are subjected to quality control to assure product quality and reproducibility. IPG strips are purchased dehydrated and frozen to minimize damage during handling,. When IPG strips are ready to be used, they must be rehydrated and sample must be loaded into them, before applying a voltage and focusing the proteins. The various rehydration and sample application methods are described in detail in Chapter 4. Proteins are usually focused using a stepwise voltage program until the proteins reach the steady-state. Protein samples from different sources typically require their own voltage program, described as the total number of volts applied over a given amount of time (known as Volt-hours, Vh).

New Reagents for 2-DE

In addition to the development of IPG strips, many new reagents have been created, or adapted, for use in IEF (Fountoulakis et al, 2001). One of the major trouble-spots with using IPG strips is maintaining protein solubility after cellular extraction and

during IEF. In aqueous solutions, charged proteins have their minimal solubility when their net charge becomes zero and hydrophobic proteins often have relatively low solubility even with the presence of detergent. During IEF, one normally uses a "buffer" solution to rehydrate the IPG strips and help maintain protein solubility. The IEF solution usually contains chaotropes, nonionic or zwitterionic detergents, and reducing agents. Salts are to be avoided or used at very low concentrations to minimize heating problems associated with high currents. For example, Tris-HCl or phosphate buffers can be used but their concentration should be ≤ 50 mM. The reagents work in concert with one another to denature the protein and help to maintain protein solubility. Chaotropes disrupt H-bonding and are usually urea or thiourea (Figure 14). Nonionic detergents include Nonidet 40 (NP-40), n-octylglucoside, Triton X-114 and 110. Zwitterionic detergents include CHAPS, deoxycholate, and sulfobetaine derivatives (amidosulfobetaine 14, ASB-14, and sulfobetaine 3-10, SB 3-10). Sodium dodecyl sulfate can be used but must be present at concentrations $< 0.25\%$ (Figure 15). Additionally, the chemical tributyl phosphine (TBP) is a nonionic reducing agent that can maintain a reducing environment during IEF. This is especially important for proteins with pI's between 7.5 and 10, since this is the pH range that is optimal for creating or shuffling disulfide bonds. A water-soluble phosphine, tricarboxyethyl phosphine (TCEP) has also been used as a reducing agent.

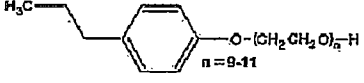
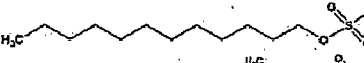
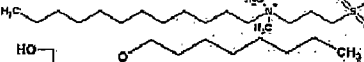
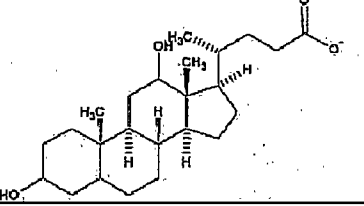
These new reagents can also be used to extract proteins from cells. Thiourea has been especially useful in recovering outer membrane, integral membrane and peripheral membrane proteins (Molloy, 2000; Rabilloud, 1996; Rabilloud et al, 1997; Rabilloud, 1998; Rabilloud, 1999). Detergents help to extract membrane-associated proteins by

coating their hydrophobic regions with detergent and creating mixed protein/detergent micelles (which act like pseudomembranes). TBP is nonionic and is able to easily penetrate membranes and detergent micelles in order to reduce disulfide bonds, thereby further inducing protein denaturation and disrupting inter- and intramolecular interactions. A review of the protocols which use various combinations of chaotrope, detergents, and reducing agents can be found by Malloy (Molloy, 2000). The properties of nondetergent sulfobetaines can be found in studies by Vuillard, et. al, (Vuillard et al, 1995; Vuillard et al, 1998) and for the zwitterionic detergents ASB-14 and SB 3-10 in a study by Chevallet (Chevallet et al, 1998).

New Reagents for 2-DE

- | | |
|----------------------------|---|
| • Thiourea | Improves membrane protein solubility during extraction and IEF.
Requires urea for solubilization |
| • Tributyl Phosphine | Nonionic, will penetrate membranes to reduce disulfides and will not migrate during IEF, helping to maintain protein solubility.
Stoichiometric binding, toxic, poss. stability issue. |
| • SB 3-10
ASB 14
C80 | More hydrophobic detergents for extracting peripheral and integral membrane proteins.
May be hard to remove during equilibration |
| • SDS | Keep <0.25% during IEF. With heating, it is a very effective protease inhibitor. |

Figure 14. Types and properties of new reagents for 2-DE.

Examples of Detergents and Their Properties			
Name	Structure	CMC	N (MW)
Nonidet® (N) P-40		0.3 mM	100-155 (647)
sodium dodecyl sulfate (SDS)		8.3 mM	62 (288)
sulfobetaine (SB12)		3.6 mM	55 (336)
n-octylglucoside		14.5 mM	20-25 (292)
deoxycholate		20 mM	3-12 (417)

A few examples of detergents illustrating the different types of hydrophilic and hydrophobic groups. Critical micellar concentrations (CMC) and aggregation numbers (N) are also shown. MW is the average molecular weight in daltons.

Figure 15. Structure and properties of detergents. ASB-14 and SB 3-10 are sulfobetaine derivatives. SDS can be used in IEF but must be present in concentrations $< 0.25\%$.

Equilibration for Second-dimension Separations

Upon completion of IEF, the strips can be used immediately or stored frozen at -20 to -80°C . To prepare focused IPG strips for the second dimension separation, they must first be "equilibrated". Equilibration is a step that exchanges IEF detergents from focused proteins with negatively charged SDS. The intrinsic negative charge placed on the protein is not a factor in second dimension separations. Rather, SDS linearizes proteins by binding 1 SDS molecule per 2 amino acid residues (or 1.4 g SDS per 1.0 g protein) and proteins are separated according to their mobility through a porous second dimension

polyacrylamide gel (i.e., according to their molecular sieving properties). Proteins are reduced and alkylated during equilibration to ensure that each protein is completely linearized with SDS, not bound to other proteins, and to reduce artifactual results due to inter- or intramolecular bonding. A typical equilibration protocol would include: 6 M Urea, 50 mM Tris-HCl, pH 8.8, 20% glycerol, 2% SDS, and 1% DTT for 15 min followed by 6M Urea, 50mM Tris-HCl, pH 8.8, 20% glycerol, 2% SDS, and 4% Iodoacetamide for 15 min.

SDS-PAGE

Once equilibrated, the IPG strip containing the focused proteins is placed atop polyacrylamide slab gels. The IPG strips are held in place by embedding them in hot (70°C) 0.5% agarose, which solidifies upon cooling. Different percentages of acrylamide in gels provide characteristic separation ranges. Single percentage acrylamide gels (also known as homogeneous or linear gels) can be used or a range of acrylamide percentages can be poured creating gradient gels (Figure 16) (Rabilloud, 2002). Most of the gels used in this study are 8-18% acrylamide yielding a separation range of 200,000-10,000 Da.

Polyacrylamide gels are formed by copolymerization of acrylamide and bis-acrylamide (N,N'-methylene-bis-acrylamide). The reaction is a vinyl addition polymerization initiated by a free radical generating system (Gorg et al, 2000). Polymerization is initiated by ammonium persulfate and catalyzed by TEMED (tetramethylethylenediamine).

Recommended Acrylamide Concentrations for Protein Separation	% Acrylamide in resolving gel	Separation size range (MW 10^{-3})
Single percentage:		
	5%	36-200
	7.5%	24-200
	10%	14-200
	12.5%	14-100[1]
	15%	14-60[1]
Gradient:		
	5-15%	14-200
	5-20%	10-200
	10-20%	10-150

[1] The larger proteins fail to move significantly into the gel.

<http://www1.amershambiosciences.com>

Figure 16. Percentage of acrylamide and corresponding protein molecular weight separation ranges. Pore size is determined by the total % acrylamide species used and the percent of that total that is bis-acrylamide.

TEMED promotes the formation of free radicals from persulfate which catalyzes the covalent linkage between acrylamide and bis-acrylamide (Figure 17). The elongated polymer is a collection of randomly cross-linked acrylamide and bis-acrylamide.

Polyacrylamide gels are described according to the total monomer concentration, %T and the weight percentage of cross-linker, %C:

$$\%T = (\text{grams acrylamide} + \text{grams cross-linker}) / \text{total volume, mL} \times 100$$

$$\%C = \text{grams cross-linker} / (\text{grams acrylamide} + \text{grams cross-linker}) \times 100.$$

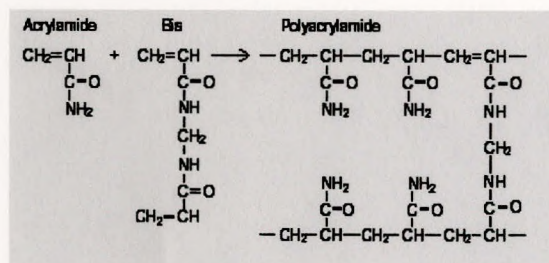


Figure 17. Polymerization of acrylamide and bis-acrylamide. This process is promoted by free radicals generated by persulfate and catalyzed by TEMED.

SDS-PAGE is usually performed using variations of the discontinuous Laemmli buffer system (Laemmli, 1970). The most common variation consists of 25 mM Tris, 192 mM Glycine, pH 8.3, and 0.1% SDS. The buffer system is referred to as “discontinuous” because the Tris-Glycine buffer is physically separated in the tank (used to perform the separation) by a polyacrylamide gel containing Tris-HCl (usually pH 8.8) and the affect focuses proteins into thin zones on top of the separation zone.. The Laemmli buffer can also contain bromophenol blue (BPB) as a tracking dye to monitor the progress of the second dimension separation. Once the IPG strips are embedded atop the second dimension gels a constant prefocusing current of ~ 5 mA per gel is used for roughly 2 h to draw the proteins out of the strip and onto the gel, then a current of ~ 20 mA per gel is used until the BPB dye front reaches the bottom of the gel.

Visualizing Proteins on 2-D Gels

There are many means for detecting proteins from 2-D gels (Neuhoff et al, 1990; Patton, 1995; Patton, 2002; Rabilloud, 2000). The method used is usually dictated by factors which include protein loading, the purpose of the gel (for protein quantitation or for blotting and chemical characterization), and the sensitivity required. Typical stains include radiolabelling with ^{35}S Met or ^{14}C , silver staining, coomassie brilliant blue dyes (CBB), Ponceau S, Amido Black, and the fluorescent stain, SYPRO Ruby. Protein spots are detected using a densitometer, fluorescent gel-scanner, chemiluminescence detector, or with radioisotopic imaging, depending on how the proteins were treated. Although most

means give some indication of the quantities of protein present, in general most cannot be used for global quantitation. The exceptions include radiolabelling methods and the use of the fluorescent stain, SYPRO Ruby. This is because other stains have inconsistent staining properties over a wide range of protein concentrations, isoelectric points and amino acid compositions, and types of post-translational modifications, such as glycosylation (Krishna et al, 1993). Furthermore there are large differences observed in the staining pattern of a gel when identical gels are subjected to different stains which indicates that different noncovalent binding mechanisms exist among the stains. The only stains currently available that have similar staining properties are SYPRO Ruby and CBB. These stains are complimentary in that fluorescently stained gels can be imaged using a commercially available scanner in order to obtain high resolution images (which leads to more accurate and reproducible spot quantitation) and then the gels can be restained with CBB to visually see the spots for eventual spot picking. Both SYPRO Ruby and colloidal CBB were used throughout this study, since SYPRO Ruby provided the required sensitivity and the widest linear dynamic range (1000-fold), CBB could be used to visualize most all of the proteins imaged in SYPRO Ruby stained gels, and we could avoid having to use radioisotopes (see Chapter 4 for more details). A comparison of the sensitivities of some of the more common stains is shown in Table 3.

Image Analysis Software

Once the protein spots have been detected and imaged on the gels, one must somehow distinguish those proteins which are differentially expressed, so that they may be plucked from the gel and further analyzed for identification. The human eye is good at discriminating spot differences in gels but not so good at deciphering quantitative differences. Thankfully, suitable image analysis software has been developed and has been one of those tools that continues to evolve. There is a growing list of available software packages that are available, each having their own features which give them advantages and disadvantages over each other. Two recent investigations have been published comparing the capabilities of some of the commercially available software packages (Nishihara et al, 2002; Raman et al, 2002). In those studies, comparative analyses of spot detection, gel matching, and spot quantity variability were reported for Melanie 3 (GeneBio) and Z3 (Compugen) (Nishihara and Champion, 2002); while sensitivity for spot detection, spot quantity and spot reproducibility were reported for Z3, PDQuest (BioRad), and Phoretix (Nonlinear Dynamics) (Raman et al, 2002). In the latter study, all three programs produced similar correlation of variations in spot detection (4-11%) and spot quantitation (3-33%), but PDQuest and Phoretix were found more suitable for examining individual spot quantities across a series of gels (Raman et al, 2002). I have also evaluated the capabilities of these and other software packages and believe that PDQuest is the one software program currently available that is sophisticated enough to perform the types of analyses needed, and yet flexible enough to allow manipulation of the

parameters (used to perform those analyses) in a user-friendly fashion. Thus, we have used PDQuest for all of our gel analysis work. (For a explanation of how PDQuest functions, see Chapter 4).

The Need for Mass Spectrometry

Traditional methods to determine protein purity, molecular weight, and identity have involved a combination of techniques such as electrophoresis, ion-exchange chromatography, size exclusion chromatography, reversed phase high performance liquid chromatography, isoelectric focusing, SDS-PAGE, and ultracentrifugation. These techniques, however, are limited by their mass accuracy and inability to detect impurities below 5% (Carr et al, 1991). For example, SDS-PAGE separates biomolecules based on molecular size and it is often used to assign an approximate molecular mass to a biomolecule based on its migration distance in a gel. It fails, however, to take into account the differences that may exist between a molecule's actual mass and its migration mass (Jardine, 1990). It is precisely this difference that restricts the mass accuracy of SDS-PAGE to 5-10%, in ideal cases and worse in others. Furthermore, all of the traditional techniques are limited in their ability to detect small mass differences between molecular species (Carr et al, 1991; Jardine, 1990). This limitation is particularly important to proteomic scientists, who wish to monitor the presence of post-translational modifications, such as acetylation, oxidation, deamidation, phosphorylation, disulfide bond formation, and many others (Sickmann et al, 2001). For these, high mass accuracy is

required. In the not too distant past, scientists needing amino acid sequence information had to rely solely on chemical methods, such as Edman degradation, to sequence proteins and peptides. Chemical methods, however, suffer from difficulties with blocked N-termini, compatibility issues with detergents and chaotropes, and required relatively large amounts of samples (Rose et al, 1984). Hence, alternative methods for protein and peptide characterizations were needed, and mass spectrometry was adapted to fill the void.

Once a protein spot has been harvested from a gel, its content can be characterized using a variety of well established MS methods. Molecular mass measurements can routinely be determined to an accuracy of 0.1%-0.01% (as compared to 5-10% with 2-D gels) (Jardine, 1990). Protein sequence information can be revealed by mapping the resulting peptides from chemical or proteolytic digestion (see peptide mass fingerprinting section below). If a tandem mass spectrometer (MS/MS) is available, further confirmation of the identity of peptide ions can be obtained from fragment ions to reveal direct amino acid sequence information (Loo et al, 1999; Ogorzalek et al, 1997).

MALDI TOF MS

The advent of matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) and electrospray ionization mass spectrometry (ESI-MS) has greatly enhanced biochemical research. These are soft ionization techniques that have been used to study protein folding and dynamics, protein structure and function, and noncovalent interactions in protein complexes, etc. (Andersen et al, 1996; Burlingame et al, 1990;

Nguyen et al, 1995; Roepstorff, 2000). MALDI-MS mixes analytes in a laser-absorbing matrix and desorbs the analytes into the gas phase, using a nanosecond laser flash, then accelerates the analytes into a time of flight (TOF) mass analyzer region (Chait et al, 1992; Hillenkamp et al, 1991). This ionization method has become popular because of MALDI's ability to ionize proteins and peptides in the presence of detergents and salts, reagents commonly used to solubilize and isolate cellular proteins. ESI-MS requires the formation of highly charged droplets which shrink due to solvent evaporation and charge repulsion thereby emitting molecular ions which are carried into the mass analyzer region (Chait and Kent, 1992; Fenn et al, 1989).

Non 2-DE Proteomic Approaches

Non-2-DE proteomic methods were not used in this study but are briefly described for a comparative perspective. Alternative approaches exist which take advantage of the physiochemical properties of proteins and peptides. All of the techniques share the strategy of separating proteins in complementary ways, such as charge and size.

Ion-Exchange Chromatography

Ion-exchange chromatography (IEC) was the first type of chromatography to be developed into a high-performance mode (Henry, 1990). Major advances were made in the 1970's with the arrival of silica-based ion-exchange resins. IEC has recently

reemerged as an excellent way to separate (and purify) proteins according to their charge. While the resolution of this technique is less than isoelectric focusing, it is more amenable to automated two-dimensional liquid separations, such as IEC followed by RP HPLC (see below for details). (Recall, that proteins and peptides are amphoteric molecules, having containing charged moieties and hydrophobic regions). Proteins are positively charged at a pH which is less than their pI and negatively charged at pH's above their pI. IEC affects the separation of peptides and proteins by taking advantage of these pH-dependent charges. Specifically, the mechanism of separation in IEC depends primarily on reversible ionic interactions between charged solutes (including salts) in the mobile phase and a charged ion-exchange group contained on the stationary phase.

There are two main types of IEC: anion exchange and cation-exchange. In anion exchange, negatively charged peptides or proteins compete with negatively charged mobile-phase ions for the positive groups exposed on the surface of the beads within the column. In contrast, cation-exchange involves competition between positively charged molecules in the mobile phase for a negatively charged stationary phase group. Generally, the stationary phase of the column is specifically made containing functional groups which can be positively or negatively charged before any separation begins.

IEC occurs in multiple steps: including the movement of solute from the mobile phase into the stationary phase via an electrostatic interaction and the selective displacement and elution of the solute (Henry, 1990). Separation occurs because each charged solute interacts differently to the charged stationary phase, via weak electrostatic interactions (Burke et al, 1989). For example in cation-exchange chromatography, the

column is negatively charged and in equilibrium with the positively charged counterions (such as the quaternary ammonium ion) present in the mobile phase. Under low pH/low counterion conditions, positively charged proteins or peptides initially displace the few positively charged counterions attached to the stationary phase. Then to selectively elute solute from the column, the concentration of counterions (NH_4^+) is slowly increased in a gradient fashion. This increase creates a competition between the counterion and the solute for the negatively charged stationary phase. Solutes with fewer positively charged moieties have fewer electrostatic interactions and are more easily displaced. Thus, solutes are eluted in order of increasing positive moieties (charge).

Reversed Phase High Performance Liquid Chromatography

High performance liquid chromatography (HPLC) describes the use of small, uniform, and rigid supports to provide efficient separations at fast flow rates and high pressures. Reversed phase HPLC (RP-HPLC) employs a nonpolar stationary phase commonly consisting of alkyl groups bonded to silica supports. Some of the more common alkyl groups include n-octadecyl (C_{18}), n-octyl (C_8), n-butyl (C_4), and phenyl. This technique separates molecules based on differences in their hydrophobic interactions with the stationary phase. An aqueous mobile phase, which is more polar than the stationary phase, is selected to deposit relatively nonpolar molecules onto the column, while the polar molecules remain in the mobile phase. A second mobile phase, is selected so that it provides similar intermolecular interactions as the stationary phase, and is used to

elute solutes from the column. The second phase is normally an organic solvent such as methanol or acetonitrile. In general, solutes elute in the order of increasing hydrophobicity when a continuous gradient is used which increases the percentage of organic phase flowing through the column. This technique has enjoyed considerable popularity because of its versatility, reproducibility, and fast analysis time. However, problems still exist in achieving high efficiency separations (as seen in GC separations) because of the inability to control mobile phase mass transfer to the stagnant mobile phase and stationary phase within HPLC columns (Mant et al, 2002; Wilce et al, 1993; Wirth et al, 1991).

RP-HPLC separations can also be influenced by using ionic modifiers and surfactants in the mobile phase, and by adjusting the pH and temperature of the mobile phases. When silica based columns are derivatized to contain alkyl groups, some silica remains unbound. This creates sites within the column that are negatively charged at pH >7. Thus, if the mobile phase pH is > pH 7, basic biomolecules can strongly interact with the exposed silica. To minimize ionic interactions and maximize hydrophobic interactions low pH solutions (pH 2-6) are used. Low pH solutions neutralize exposed silanol groups. However, low pH solutions can also positively charge solutes, like proteins and peptides. To counter the pH affect on solutes, a popular ionic modifier called trifluoroacetic acid (TFA) is used. (A variety of ionic modifiers have been used. Formic acid (FA) is often substituted for TFA but its mechanism is to enhance the solubility of hydrophobic molecules rather than act as an ion pairing agent and is better suited for nonprotein/peptide RP-HPLC separations.) TFA acts as an ion pairing agent by electrostatically interacting with the charged groups on proteins and peptides and neutralizing them, allowing for

higher resolution separations. Additionally, modifiers such as TFA or FA yield an added benefit by creating an environment which creates positively charged molecules (via the low pH conditions) which ultimately improves their gas phase partitioning and detection in mass spectrometry (when using the positive ion detection mode).

The mechanism of peptide and protein retention is complex, but studies have revealed that there is a narrowly defined concentration of organic phase that will elute the analytes off RP columns (Mant and Hodges, 2002; Wilce et al, 1993; Wirth et al, 1991). In general, peptides require a lesser mole fraction of organic solvent compared to proteins; thus, they have shorter retention times than proteins, under the same gradient conditions. Furthermore, the size of proteins limits them to absorption/desorption events, while peptides tend to partition between the mobile phase and stationary phase, undergoing thousands of interactions and improving efficiency (Hodges et al, 1988).

The evolution in column technology has created column packings that are extremely robust and reproducible. Columns are now available with a variety of internal diameters, lengths, and packing materials. Further, HPLC systems can be interfaced with mass spectrometers for "on-line" chromatographic separations followed by molecular weight determinations (for example, LC-ESI MS/MS). For maximum sensitivity, a microbore (1.0 mm x 5 cm) or capillary column (0.050 mm x 5cm) can be used.

Size Exclusion Chromatography

One technique that is not normally mentioned as a proteomic tool is size exclusion chromatography (SEC). SEC, however, is a method that has two very important characteristics applicable to proteomic studies: 1) it can be used to help purify multiprotein complexes and 2) it can tell us the approximate molecular weight of those complexes. Once a complex is isolated, further manipulation can tell us the identity of the proteins involved and give insight into *in vivo* protein-protein interactions. SEC is a liquid chromatography method whereby macromolecules are separated on the basis of molecular size or hydrodynamic radius, as they move through a column (Hodges et al, 1988; Josic et al, 1998; Mant et al, 1987). The degree of separation is determined by interactions with the pores of the stationary phase within the column. Smaller molecules interact with the pores more readily than larger molecules. As a result, larger molecules elute from the column first followed by the smaller molecules (just the reverse from SDS-PAGE where small molecules migrate through the gel faster than larger molecules).

Stationary phases for aqueous SEC are usually cross-linked gels or derivatized silica. Silica bead columns offer the advantage of mechanical strength and uniform particle size but are relatively expensive. Silica based columns are used in cases that require high efficiency or high flow rates. Cross-linked gels, on the other hand, offer gels which do not contain any underivatized silanol groups, are inexpensive, but yield low resolution. Recall, that charged silanols can interfere with protein and protein separations.

A wide variety of cross-linked gels have been used as stationary phases to separate proteins. Some of the more common types include: Sephadex, a semi-compressible cross-linked dextran; Sepharose, a compressible cross-linked agarose; Bio-Gel A, a compressible agarose; Superose, a rigid cross-linked agarose; and Sephacryl, a semi-compressible alkyl dextran cross-linked with N, N'-methylene bisacrylamide (advances in chrom). The selection of which gel to use depends upon the range of molecular weight separation desired, flowrate, and resolution needed.

Electrospray Mass Spectrometry

A popular tool used for the characterization of the structure, sequence, and purity of proteins and peptides is electrospray ionization mass spectrometry (ESI MS). Both MALDI and ESI MS have become the quintessential tools for the analysis of many biomolecules.

A valuable aspect of ESI MS is the creation of multiply charged protein ions during the electrospray process (Burlingame et al, 1990; Fenn et al, 1989). The electrospray spectra of proteins are therefore characterized by a coherent sequence of protonated molecular ion peaks with different numbers of charges. Adjacent peaks in a sequence differ by only one charge, which is almost always a proton for peptides and proteins (Figure 18). By solving for two equations with two unknowns the spectra can be deconvoluted to yield the intact mass of the protein or peptide (Figure 19). Typically, the

resolution of ESI mass spectrometers is enough to tentatively identify post-translational modifications.

As a direct result of multiple charging, mass spectrometers with a limited mass-to-charge range are able to be used in the analysis of proteins, because a protein's multiply charged distribution falls within that range. Charging is generally consistent with the number of basic residues (Lys, His, Arg) and exposed N-terminal amine, under acidic conditions. However, the conformation of a protein, solution composition, and solvent pH can also influence the degree of charging. For example, tightly bound proteins may undergo limited charging because the basic residues are not exposed; but acidic solution conditions or heating may aid in denaturing the protein enough to allow sufficient charging. Some classes of proteins have remained difficult to analyze via ESI MS. These include very hydrophobic proteins, insoluble membrane proteins, and proteins with very few basic sites. Nevertheless, ESI MS has rapidly become widely used for the study of proteins from 10 to 70 kDa and ESI MS/MS the method of choice for peptide sequencing.

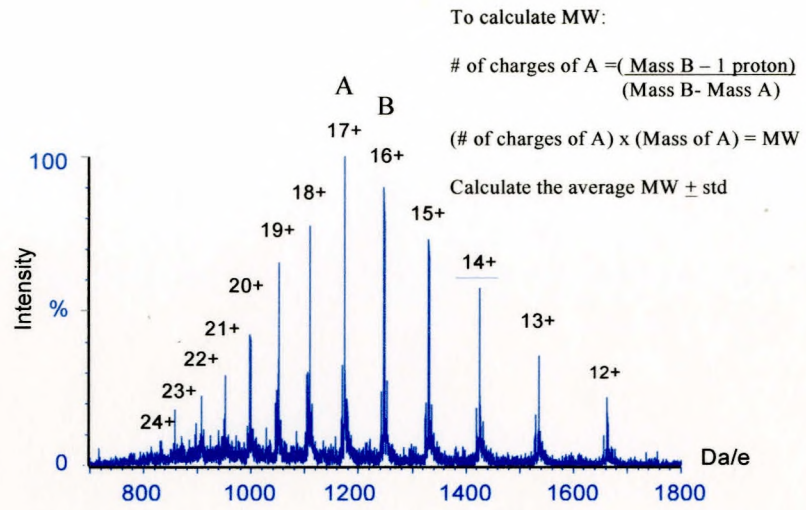


Figure 18. Example of a multiply charged series for a protein.

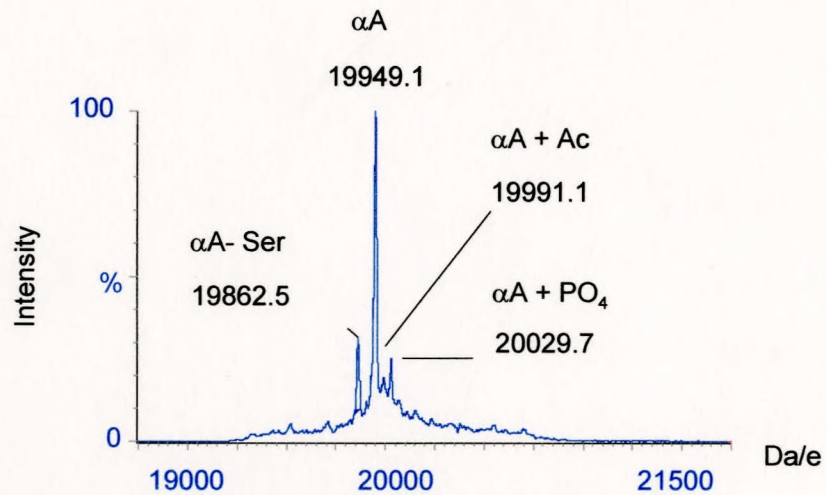


Figure 19. Deconvolution of the multiply charged series in figure 14 yields the intact molecular weight of this protein (α A-crystallin lens protein).

The process of electrospray is commonly referred to as the electrostatic dispersion of a liquid. Generally, a high electric field is applied between a liquid flowing through a capillary and a sampling cone counter electrode. The internal diameter of the capillary, the flow rate of the liquid, the presence of a conducting liquid, and the electric field work together to form highly charged droplets of a limited size. A nebulizing gas, dry bath gas, and heat are then used in conjunction with the charged droplets to aid in solvent evaporation and the stable production of molecular ions. Typical flow rates are from 1-10 $\mu\text{L}/\text{min}$ through capillaries of 50-100 μM and electric field strengths of 2-5 kV are commonly used. If the field strength is too high the spray will become unstable as the electric field at the tip of the capillary is disrupted by the onset of corona discharge; if too low, inefficient charging will drastically decrease ion efficiency. There are two basic mechanisms commonly used to explain the transfer of ions from solution to the gas phase. The first mechanism was proposed by Malcolm Dole and requires the formation of very small droplets containing a single ion; this is sometimes referred to as the single ion droplet theory (SIDT) (Fenn et al, 1989). Dole proposed that the highly charged droplets shrink by solvent evaporation and coulombic explosions until nanometer sized droplets are formed containing a single ion. The solvent molecules then evaporate and leave a single molecular ion. The other was proposed by Iribarne and Thomson and is known as the field evaporation model (Gamero-Castano et al, 2000). In this model, charged droplets undergo repeated solvent evaporation and Coulombic explosions until the ions are field emitted from small, highly charged droplets. The difference between the two models is that the

SIDT theory is a passive process while the ionization process itself plays an active role in the field evaporation model.

Currently, many different commercial instruments are available that reliably produce ions by electrospray. Two of the most popular instruments for protein and peptide analysis by mass spectrometry are the quadrupole and ion trap mass analyzers. In the present study, neither tool was used for the analysis of proteins or peptides isolated from 2-D gels, however, such instruments are briefly described below and are available for use here at MSU. Quadrupoles and ion traps typically use a hexapole lens for focusing and/or a short quadrupole for mass prefiltering, along with their corresponding quadrupole or trap for mass filtering. Most instruments also use an inlet device which enhances desolvation and ionization efficiency, such as the heated inlet capillary on the LCQ or Esquire 3000. The potential between the probe and the sampling cone is 3.5 kV for conventional flowrates (5 μ L/min) and much lower the slower the flowrate. Typically, a liquid sample is carried into a 75 μ M i.d. stainless steel capillary at a flow rate of 5 μ L/min from which a nitrogen sheath gas nebulizes highly charged droplets into the heated inlet and carries the desolvated molecular ions into the mass analyzer region for eventual detection (Jonscher et al, 1997).

Peptide Mass Fingerprinting and Protein Identification

Techniques for the identification of proteins by peptide mass, known as peptide mass fingerprinting (PMF), have been described in detail by Pappin and Mann (Aebersold et al, 2003; Pappin, 1997; Pappin, 2003; Park et al, 2001; Roepstorff, 2000; Soskic et al, 2001; Stensballe et al, 2001; Yates, 1998), among others. The technique involves the generation of peptides from proteins usually using residue-specific enzymes, the determination of peptide masses, and the matching of these masses to theoretical peptide libraries generated from protein sequence databases. As proteins have different amino acid sequences, their peptides generally produce characteristic “fingerprints”.

The first step of peptide mass fingerprinting is protein digestion. Proteins are usually enzymatically digested *in situ* within a gel matrix. The enzyme of choice is currently trypsin (which cleaves specifically at lysines and arginines), but other enzymes like Lys-C or V8 have been used (Pappin, 1997). Chemical digestion by cyanogen bromide (which cleaves specifically at methionine residues) has only recently been used and preliminary results indicate that it is a viable and efficient option, especially for 2-DE separated membrane proteins. Some protocols reduce and alkylate the proteins within the gel piece before performing proteolysis, however, no solid evidence exists supporting the need to do so. Besides, most protocols used for 2-DE reduce and alkylate IEF separated proteins during the equilibration step and before SDS-PAGE. There is a possibility, nonetheless, that some proteins within the strip may not get fully reduced and alkylated. Further, even though proteins are separated under denaturing conditions using SDS-

PAGE, the removal of detergent is required before staining with coomassie dyes or fluorescent stains. Once the detergent is removed, any proteins containing uncapped thio-groups may form intramolecular disulfide bonds and subsequently precipitate or obtain a three dimensional structure which makes proteolysis less efficient. Most staining/destaining protocols, however, are performed under acidic conditions and it is usually during these steps that one has to worry about disulfide formation. It is, therefore, very important to use pH conditions or buffers which limit disulfide formation or use a protocol which incorporates a reduction/alkylation step.

After proteins are digested, peptide masses are determined by mass spectrometry. Direct analysis of peptide mixtures can be achieved by ESI MS or MALDI TOF MS. (For an in-gel digestion and extraction protocol see Chapter 5.) Direct analysis by MALDI MS is preferable because of its higher sensitivity and greater tolerance to contaminating substances from 2-D gels (Pappin, 2003; Stensballe and Jensen, 2001). However, one must exercise good judgement when identifying proteins purely by the number and "closeness" of peptide matches. A major challenge associated with PMF (which strictly uses peptide mass information) is data interpretation prior and subsequent to computer matching against libraries of theoretical peptide digests. Spectra must be carefully examined to determine which peaks represent peptide masses of interests, as there are often multiple proteins present in the digest, and contaminating substances such as enzyme autodigestion products and keratin. Some of these interfering peaks can be removed before database searching by using the mass spectrometry software to subtract out a predetermined mass-list of contaminants.

A number of computer programs are now available for matching peptide masses against databases. Matching is usually best undertaken with human interaction, whereby peaks of mass 800-3000 m/z are matched under selected search parameters. These parameters include incorporating any covalent modifications known to be present on the proteins (such as carboxyamidomethylation on cysteines which occurred during alkylation, using iodoacetamide), any variable modifications (such as oxidation on methionines or phosphorylation on serines and threonines), and the maximum mass difference to allow for between the experimental mass and the theoretical mass (Kovtoun et al, 2002; Soskic and Godovac-Zimmermann, 2001; Stensballe and Jensen, 2001). The latter parameter is the most important one for determining whether a database matching result is correct or not. High mass accuracy is the key to obtaining unambiguous database results.

The best way to obtain highly accurate m/z values for peptides using MALDI MS is to use internal calibrants during data acquisition. One may spike each sample with a calibrant(s) such as angiotensin or use "contaminant" peaks already present in the sample, such as trypsin autolysis fragments or even keratin fragments. The most reproducible and least labor intensive method is to use trypsin fragments for calibration. However, since MALDI MS is susceptible to ionization suppression effects, one must be careful not to use too high a concentration of trypsin or one may not detect any peptides of interest. To limit the amount of trypsin autodigestion, methyl-modified trypsin is used. This trypsin has been reductively-methylated at lysines and is not as prone to autodigestion as the unmodified form. Autodigestion of trypsin can still occur at arginines, so we may use

specific autodigestion fragments as internal calibrants. A complete protocol for in-gel trypsin digestion can be found in Chapter 5 or in the mass spectrometry user's guide.

One of the bottle-necks in using the PMF approach is interpreting the database search results. Many different variables and searching parameters can cause false positive matches, most notably of which include: improperly calibrated spectra, matching peptide masses to contaminants, assuming too many missed cleavages occurred, using too many variable modifications, improper (presence of acrylamide adducts) or non-existent alkylation of cysteines, and allowing mass errors > 200 ppm. The best way to obtain the most accurate results is to manually evaluate each database result as to its validity. New software algorithms are desperately needed and are currently limiting the utility of the PMF approach.

An extension of peptide mass fingerprinting, called peptide sequence tagging, is also a popular method to identify proteins (Aebersold and Mann, 2003; Yates, 1998). This technique uses tandem mass spectrometry (MS/MS) to initially determine the mass of each peptide, then subjects each of them to a collision with gas which causes the peptide to fragment. The fragments from the resulting spectra gives information about each peptide's amino acid sequence and the information exported along with the corresponding intact peptide mass for database matching. Protein identifications made in this fashion contain twice the evidence as peptide mass fingerprinting alone. The complimentary nature of the two techniques allows for the identification of the amino acid residue which is post-translationally modified. Further, protocols can be used in preparing for PMF that are compatible for ESI- MS/MS, such as the one used for protein spots harvested from gels

(see Chapter 5). The drawbacks to this technique are the complexity of the data generated, the level of expertise required for data interpretation, and the skills required to operate MS/MS instruments. Nevertheless, this technique represents a powerful and complimentary or alternative approach to identifying proteins from PMF (Mann et al, 2003).

To date, there are only a handful of publications that have used the power of proteomics to isolate and characterize proteins from extremophiles. Most research has involved the cloning and expression of genes or gene fragments in *Escherichia coli* in order to determine gene function (see Chapter 5). While this is a very clever way to gain valuable insight into the function of gene and specifically expressed proteins in species that are hard to grow in the lab, the larger questions about which proteins are being expressed under the conditions used, the extent each protein naturally expressed, and the relationship the protein has to other proteins can only be arrived at by conjecture. Proteomic applications, on the other hand, using the tools and techniques discussed above will allow us to directly study the dynamic relationships between proteins, which ultimately control and maintain cellular processes.

CHAPTER 4

QUANTITATIVE EVALUATION OF SAMPLE APPLICATION METHODS FOR
SEMI-PREPARATIVE SEPARATIONS OF BASIC PROTEINS BY 2-DEChapter Summary

The use of cup-loading for sample application has become widely used in 2-DE for resolution of basic proteins, but no side by side quantitative study has been published which compares cup-loading with the alternative passive and active rehydration methods to fully promote one type of loading method over another. Replicate 2-D gels from each loading method were quantitatively evaluated for gel-to-gel reproducibility using IPG 6-11 strips and semi-preparative protein loads (300 μ g). Gels were stained with SYPRO Ruby and analyzed with PDQuest. An inexpensive homemade assembly for cup-loading was used with the Protean IEF Cell for separation of whole cell extracts from the archaeon, *Sulfolobus solfataricus*. Cup-loading was determined to be far superior for IPG 6-11 separations than active or passive rehydration methods. Cup-loading consistently produced the greatest number of detectable spots, the best spot matching efficiency (56%), lowest spot quantity variations (28% CV), and the best looking gels qualitatively. The least satisfactory results were obtained with active rehydration, followed closely by passive rehydration in off-line tubes. Passive rehydration experiments, performed using an on-line IEF tray, produced comparable spot numbers to cup-loading (84%), with 55 of the spots having higher abundance but 10% more spot quantity variance than cup-loading.

Introduction

A central goal in proteomics is to elicit the maximum amount of information on changes in protein expression patterns in response to biological stimuli to globally analyze complex biological systems. This analysis requires the ability to separate and resolve the largest number of cellular proteins, track protein expression levels, identify proteins of interest, and assign protein function where possible, using bioinformatics. In classical proteomics, the linchpin is producing high quality, reproducible gels with high resolution. It is only after high quality gels are obtained that image analysis software can be used to precisely identify changes in the patterns of protein expression. In order to produce the best gels many factors must be considered, including sample preparation methods, the isoelectric focusing range, the percent acrylamide in the second-dimension gel, the type of protein staining used, and the detection capabilities of the hardware employed for imaging. Equally important is the way in which the sample is applied to an IPG strip and the amount of protein loaded.

The impact of different sample loading techniques on gel quality has received considerable attention over the years. Most investigations have been qualitative in nature, however, and while very valuable, lack the quantitative information necessary to fully promote one sample loading method over another. Specifically, studies that compare different sample loading techniques in terms of gel-to-gel reproducibility, spot quantity variance, and whether or not a particular method is biased towards certain types of proteins have been limited in scope. Zuo and Speicher recently evaluated protein

recovery during each step of 2-DE using cup-loading and passive rehydration sample applications for separating radiolabeled tris-soluble *Escherichia coli* extracts on the Multiphor and IPGphor IEF systems (Zuo et al, 2000). They did not, however, report on gel-to-gel reproducibility or provide spot quantity values and limited their study to separations over IPG 3-10. Nishihara and Champion, on the other hand, addressed the reproducibility and selected spot quantity variations of replicate 2-D gels from whole cell *E. coli* extracts, but limited their study to sample application by passive rehydration and separations over IPG 3-10 (Nishihara and Champion, 2002).

Until recently, quantifying proteins in polyacrylamide gels over a wide dynamic range, as is typically found in biological systems, has been difficult and often imprecise. One has had to either use radiometric labeling (Link, 1999) (which has safety and waste disadvantages) or accept the idiosyncrasies of various staining methods (Dunn et al, 1994; Voss et al, 2000). Many previous articles have addressed this issue and have thoroughly covered the limitations of the more common staining protocols (Brush, 1998; Dunn and Crisp, 1994; Dunn, 1997; Lopez et al, 2000; Mortz et al, 2001; Patton, 2000; Rabilloud et al, 1994; Rabilloud, 2000). Briefly, colloidal CBB staining can provide reproducible quantitation over a 20-fold linear dynamic range (50-1000 ng) and detection limits to 5 ng protein/band (Berggren et al, 2000; Brush, 1998; Neuhoff et al, 1988; Neuhoff et al, 1990; Nishihara and Champion, 2002; Patton, 2000). Optimized acidic silver nitrate staining protocols are highly sensitive with detection limits to 0.1 ng protein/band and a 50-fold linear dynamic range (1-50 ng), but are well known to suffer from inhomogeneous protein staining (Nishihara and Champion, 2002; Rabilloud, 1990;

Scheler et al, 1998; Switzer et al, 1979). Alkaline silver staining methods such as silver diamine and ammonical silver also present mixed blessings. Silver diamine staining can yield detection limits which are intermediate to CBB and acidic silver nitrate while ammonical silver staining can be just as sensitive as silver nitrate; however, alkaline methods have found limited use because of poor gel-to-gel reproducibility (Dunn and Crisp, 1994; Patton, 2000). On the other hand, staining proteins with the noncovalent fluorescent dye, SYPRO Ruby (Molecular Probes), has been shown to be quite reproducible (with similar binding characteristics as CBB), very sensitive with detection limits in the low nanogram range, compatible with downstream protein analysis methods, and to have a 1000-fold dynamic range (1-1000 ng) (Berggren et al, 2000; Nishihara and Champion, 2002; Patton, 2000).

Along with improved staining protocols, an ever-evolving selection of image analysis software has become available for spot detection and quantification. Such software is proving itself to be extremely valuable for discriminating differences in protein abundance in biological studies (Alaiya et al, 2001; Cheon et al, 2001; Cole et al, 2000; Corbett et al, 1995; Easty et al, 1991; Kim et al, 2002; Minowa et al, 2000; Salmi et al, 2002; Schwertz et al, 2002). Two recent investigations have been published comparing the capabilities of some of the commercially available software packages (Nishihara and Champion, 2002; Raman, Cheung, and Marten, 2002). In those studies, comparative analyses of spot detection, gel matching, and spot quantity variability were reported for Melanie 3 (GeneBio) and Z3 (Compugen) (Raman et al, 2002); while sensitivity for spot detection, spot quantity and spot reproducibility were reported for Z3,

PDQuest (BioRad), and Phoretix (Nonlinear Dynamics) (Nishihara and Champion, 2002).

In the latter study, all three programs produced similar correlation of variations in spot detection (4-11%) and spot quantitation (3-33%), but PDQuest and Phoretix were found more suitable for examining individual spot quantities across a series of gels (Nishihara and Champion, 2002). In the present study, we use SYPRO Ruby for fluorescent staining and PDQuest for spot detection and quantitation in order to evaluate the influence of different sample application methods on 2-D gel quality and reproducibility.

There are three different sample application methods commonly used in 2-DE: active rehydration, passive rehydration, and cup-loading. Active and passive rehydration mix protein extracts with rehydration buffer and allow a dehydrated IPG strip to rehydrate with or without the presence of an applied voltage. If no voltage is applied, proteins are passively absorbed into the strip during reswelling, whereas if a potential of 50-150 V is applied proteins are actively (as well as passively) drawn into the rehydrating strip. Cup-loading involves applying the sample within a discrete zone to a previously rehydrated strip. Cup-loading applications are usually performed at the cathodic or anodic end of an IPG strip. As a matter of convenience, sample application during rehydration is preferred and is now commonly employed when separating proteins over an acidic pH range (e.g. pH 4-7) or over a broad pH range (e.g. pH 3-10) (Rabilloud et al, 1994b; Sanchez et al, 1999). On the other hand, when focusing proteins over the more basic pH regions (e.g. pH 6-11) many investigators have demonstrated that qualitatively better separations can be obtained if the sample is applied using the cup-loading method (Gorg, 1999; Gorg et al, 2000).

In this study, we evaluate sample application by cup-loading with active and passive rehydration methods in terms of reproducibility among replicate 2-D gels for a given method, as well as between replicate gels from each method. Using a complex mixture of cytosolic and membrane proteins, extracted from whole cell lysates of *Sulfolobus solfataricus* P2 (an acidophilic and hyperthermophilic archaeon), our goal was to systematically and quantitatively evaluate different sample loading techniques for resolving the more basic proteins of a mixture, using IPG 6-11 strips. We then compared the reproducibility of the IPG 6-11 results to separating the acidic components of the same mixture over IPG 4-7's, using the active rehydration method. A semi-preparative loading of 300 µg was used in all experiments to enhance our ability to view more of the proteome over the shorter pH ranges used. The number of spots detected in each experiment, the reproducibility of each sample loading technique, the gel-to-gel matching efficiency, and the reproducibility of spot quantity within each technique was evaluated in order to determine which sample loading method is optimum for separating the basic protein components of a complex mixture.

Electrophoresis Equipment and Chemicals

The Protean® IEF Cell for isoelectric focusing and Protean® II XL Multi-Cell for SDS-PAGE, Bio-Lyte® carrier ampholytes (3/10, 5/7, 8/10), Bio-Rad Bradford protein assay, SYPRO Ruby, bromophenol blue, mineral oil, dithiothreitol, low-m agarose, tris(hydroxymethyl)-aminomethane, glycerol, 40% acrylamide/bis solution (37.5:1, 2.6

%C), ammonium persulfate, TEMED, and all wicks (cut from filter paper, catalog number 1650921) were purchased from Bio-Rad (Hercules, CA, USA). IPG strips (18 cm 4-7 and 6-11) and Immobiline sample cups were purchased from Amersham Pharmacia Biotech AB (Uppsala, Sweden). Urea (SigmaUltra), thiourea, CHAPS, tributylphosphine, iodoacetamide (SigmaUltra), lauryl sulfate (sodium salt, 99%), deoxyribonuclease I (DN-25), ribonuclease (R-5000), phenylmethylsulfonyl fluoride (PMSF), 2 mL plastic serological pipettes, and bovine serum albumin were purchased from Sigma (St. Louis, MO, USA). HPLC-grade methanol and glacial acetic acid were purchased from Fisher Scientific. Amidosulfobetaine (ASB-14) was purchased from CALBIOCHEM.

Sample Preparation

Cell Culture and Harvesting

Sulfolobus solfataricus P2 was grown aerobically at 80°C and pH 3.2 in modified 9K medium (Schleper et al, 1994) without additional vitamins and containing (wt/v) 0.1% yeast extract and 0.2% sucrose. Cultures were grown in long neck Erlenmeyer flasks in a shaking oil bath incubator for five to seven days to a late log-phase density of 0.6 (OD₆₅₀). Cells were isolated by centrifugation at 2000 X g at 25°C for 15 min and the cell pellets resuspended and washed twice in 50 mM Tris-HCl buffer, pH 7.4. The washed pellets were lyophilized overnight.

Extraction of Proteins and Preparation for IEF

Eight separate samples, labeled A-H, were individually prepared according to the following protocol. Ten milligrams of dried cells were resuspended in 1 mL of 50 mM Tris-HCl, pH 7.4 containing 1 mM PMSF and sonicated for six cycles of 30 seconds duration with intervening cooling on ice, using a Branson Sonifier 450 with titanium probe-tip, 50% duty cycle, and power output control at 2. After sonication, the lysed cells were incubated with 100 ng of DNAase and 100 ng RNAase for 1 h at room temperature. Next, the material was spun at 2000 X g for 5 min. to remove unbroken cells and cellular debris and the supernatant collected. 500 μ L of the supernatant was then added to solid urea, thiourea, CHAPS, and ASB-14 so that the final concentration was 5 M Urea, 1.5 M Thiourea, 3% CHAPS, 1% ASB-14, and 50 mM Tris-HCl. Additionally, the solution was made to 4 mM tributylphosphine (TBP) and 0.2% carrier ampholytes 3/10 and allowed to stand for at least 2 h. The sample was then spun at 14,000 X g for 10 min. on a benchtop Eppendorf centrifuge (5415) and the supernatant used for 2-DE. Each of the individual sample preps was evaluated using the various sample loading methods and a group is defined as all of those gels obtained from a specific sample prep (A-H).

Isoelectric Focusing and Sample Application

All IEF was performed using the Protean[®] IEF Cell and its 17 cm IEF tray. In order to fit the 18 cm Amersham Pharmacia strips into the IEF tray, the strips were cut to 17.5 cm, with the IPG 6-11's cut at the basic end and the IPG 4-7's cut at the acidic end. All IPG 6-11 strips and all IPG 4-7 strips were from the same batch of IPG strips, respectively. Four different sample application techniques were used to separate whole cell protein extracts on IPG 6-11's: cup-loading, active rehydration, passive rehydration in off-line tubes, and passive rehydration in the IEF tray normally used for focusing. Samples applications to the IPG 4-7's were all with active rehydration. The same IEF buffer was used in all experiments and consisted of 5 M Urea, 1.5 M Thiourea, 3% CHAPS, 1% ASB-14, 50 mM Tris-HCl, pH 7.4, 0.2% carrier ampholytes (CA) 3/10, 0.1% CA 5/7, 0.1% CA 8/10, and 2 mM TBP. The following voltage program was used for cup-loading and passive rehydration of IPG 6-11's: 100 V constant for 2.0 h, linear ramp to 300 V over 6.0 h, linear ramp to 1000 V over 2.5 h, linear ramp to 2500 V over 2.5 h, then 3500 V constant for 18 h, and 5000 V constant for 3.5 h, for a total of 87,000 Vh. All active rehydration applications were also focused to 87,000 Vh but used the program: 50 V constant for 14 h, linear ramp to 300 V over 3.5 h, linear ramp to 1000 V over 2.5 h, linear ramp to 2500 V over 2.5 h, then 3500 V constant for 18 h, and 5000 V constant for 3.5 h. Focused IPG strips were stored at -80°C before equilibration and separation on SDS-PAGE.

Cup-loading

Cup-loading experiments were performed using a simple and inexpensive homemade adaptation of the Protean IEF Cell (Figure 20). (Please note: Bio-Rad has recently made a cup-loading tray available that is compatible with the Protean® IEF Cell, but these experiments were performed before its availability and we prefer our technique given its low cost and excellent performance.)

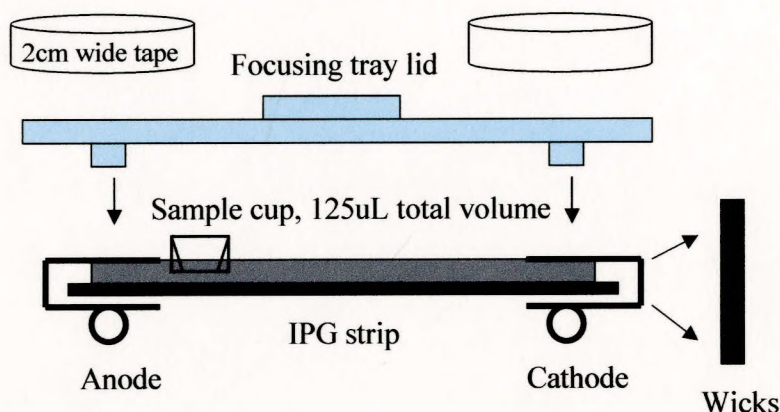


Figure 20. An inexpensive homemade configuration for cup-loading samples using the Protean IEF Cell (Bio-Rad). Filter paper wicks are cut 2.0 cm in length by 0.5 cm in width (Bio-Rad catalog no. 1650921) and used to complete the circuit from the surface of the IPG strips to the electrodes. The lid of the focusing tray is then used to press down on the wicks and hold the assembly in place.

IPG 6-11 strips were rehydrated for 24 hr with 350 μ L of IEF buffer within parafilm sealed 17.5 cm tubes (cut from 2 mL plastic serological pipets, then placed gel-side up in the lanes of the IEF tray. Long wicks were required for this homemade

adaptation and were cut 2.0 cm long by 0.5 cm wide from 0.08 cm thick filter paper, saturated in deionized 17.5 M Ω nanopure water, patted dry, and placed into the IEF cell such that one end of the wick overlapped the platinum wire electrode in the bottom of the IEF tray and the other end overlapped with the end (and gel-side) of the IPG strip. The wicks were mildly crimped at their apex, using a pair of tweezers, to better hold their position during the assembly, but the paper should not be damaged or localized heating may result. A cup is then wedged into the lane of the IEF cell and gently pushed down on top of the IPG strip, approximately 1.2 cm from the anode. Care must be taken to not push the cup too far down into the IPG strip, and for multiple cup loading experiments every other lane was used in the IEF tray. To make sure the cup is well seated, 20 μ L of IEF buffer containing bromophenol blue (BPB) was added to the cup to check for leakage. If cup leakage was apparent, the liquid was removed and the cup was gently pushed farther down into the IPG strip. After repositioning, the seating was rechecked with 20 μ L of BPB buffer. To the seated sample cup, 70 μ L of extract containing 300 μ g of protein was added along with an additional 55 μ L of IEF buffer (no BPB), so that the total volume was 125 μ L. Finally, the lid of the focusing tray was used to press down on the entire assembly and hold the wicks in place. Thus, it is critical that the wicks extend far enough over the top of the strip in order to be held down by the pegs that protrude from the bottom of the focusing-tray lid. Additionally, one can place two 2 cm wide blocks (or rolls of tape) between the upper side of the tray lid and the exterior lid of the unit to hold the assembly together, as shown in Figure 20, above.

Active Rehydration

For active rehydration experiments, 280 μL of IEF buffer was distributed evenly into a lane of the IEF tray and then 70 μL (300 μg) of sample was deposited into the IEF buffer along the length of the lane. The IPG strips were then placed gel-side down on top of the mixture and allowed to rehydrate 2 h, before adding wicks and covering with 3 mL of argon-degassed mineral oil. Homemade 0.08 cm thick wicks were cut 0.5 cm wide by 1.0 cm long, saturated with nanopure water, patted dry then placed under the IPG strip and simultaneously over the wire electrode. (When using thiourea we have found that the thicker wicks minimize precipitation near the anode better than the thinner strips and we avoid having to change the wicks during the run.)

Passive Rehydration

IEF buffer containing sample was passively loaded while reswelling IPG 6-11 strips either within off-line tubes or within the tray used for IEF. Plastic 17.5 cm length tubes, as described under cup loading, were used for off-line passive rehydration. Off-line passive rehydration is defined in this study as the in-tube method. One end of the tube was sealed with parafilm, the IPG strip was slid into the tube and 350 μL of IEF buffer, containing 70 μL (300 μg) of protein extract, was gently pipetted into the space between the gel side of the IPG strip and the inside wall of the tube. The open end of the tube was then sealed with parafilm and the tube was placed on a level surface at room temperature for 24 h. After in-tube rehydration, the strips were placed gel-side down in

the IEF tray, wicks were added, then the strips covered with 3 mL of degassed mineral oil and IEF performed. For passive rehydration using the IEF tray (defined as the in-tray method), 350 μ L of IEF buffer containing 70 μ L (300 μ g) of protein was evenly distributed in a lane of the IEF tray, the IPG strip was placed gel-side down on top of the solution and allowed to stand for 2 h, then covered with 3 mL of argon-degassed mineral oil and rehydrated for 24 h. After rehydration, wicks were inserted (no removal of oil is necessary) and IEF performed.

Equilibration and SDS-PAGE

After IEF, equilibration was performed using 6 M Urea, 50 mM Tris-HCl, pH 8.8, 30% glycerol, 4% SDS, and 2% DTT for 15 min followed by 6M Urea, 50mM Tris-HCl, pH 8.8, 30% glycerol, 4% SDS, and 8% IAA for 15 min. The IPGs were embedded in hot 0.5% agarose (70°C), containing bromophenol blue, on top of 8-18% polyacrylamide gels (20 cm x 20 cm x 1.5 mm). The second dimension gels were poured in batches and the same batch was used for all IPG strips obtained from an individual sample prep. Thus, each group of gels, representing each sample prep, was run using different batches. SDS-PAGE was performed at 5 mA per gel for 2 h, followed by 20 mA per gel until the dye front reached the bottom of the gel.

Gel Imaging

Upon completion of SDS-PAGE, gels were fixed in 10% methanol and 7% acetic acid for 1 h with gentle agitation. The fixing solution was then removed and 300 mL of SYPRO Ruby added per gel and the gels stained overnight with agitation. Gels were then destained for three hours using 10% methanol and 7% acetic acid with one change of destaining solution. Two gels were fixed and stained per plastic container. Fluorescent scanning was performed using the Bio-Rad Molecular ImagerTM FX at 200 μ M pixel resolution, using a 532 nm laser, 1064 nm blocking excitation filter, and 555 nm long-pass emission filter.

Replicates Gels and Image Analysis

Eight separate protein extractions were performed and labeled A-H. Each sample, A-D, was run in triplicate four different ways: IPG 6-11 cup-loading, IPG 6-11 active rehydration, IPG 6-11 passive rehydration (either in-tube or in-tray), and IPG 4-7 active rehydration. Each sample, E-H, was run in triplicate using active rehydration and IPG 4-7's to statistically evaluate the variance attributed to sample preparation and handling. A group is defined in these experiments as all of the replicate gels obtained from one of the sample preparations, A-H. Groups A-D were used for the four different runs described above and groups E-H were used only for active rehydration on IPG 4-7's. For a particular group, the same amount of protein was loaded on each of the strips, but the

total amount of protein loaded between groups varied slightly. Protein concentrations were based on Bradford assays using bovine serum albumin as the standard.

Spot Detection and Quantitation

Spot detection, quantitation, and analysis were performed using the PDQuest™ 2-D Analysis Software, version 7.0, from Bio-Rad. Gels were grouped according to which extract (A-H) was used and each group was processed using the same analysis parameters. Spot detection parameters were selected using the Spot Detection Wizard of PDQuest. Since the grouped gels varied in the degree of streaking, the spot detection parameters were optimized for each group and applied across all gels in that group, so that detection criteria were systematically the same across the group. Limits were put on the horizontal and vertical radius of a spot (5 and 25, respectively) to minimize the effects of streaking and maximize the number of spots detected, and these same limits were used in all analyses. Other selections used in all analyses include the floating ball method for background subtraction, a pixel size of 3 x 3 for spot detection, and the power-mean for noise filtering. We used the manual matching function instead of the automated routine for gel matching to obtain the highest gel matching. Spot detection and quantitation were determined after background subtraction and noise filtering. (More detailed information on these parameters can be found at <http://www.biorad.com>, under the literature/software search using the keyword PDQuest and in a review by W. F. Patton (Patton, 1995))

Gels were matched within a group after establishing 30-40 landmarks (or anchor

spots) for the IPG 6-11 gels and 55-100 landmarks for the IPG 4-7 gels. Landmarks were selected from all regions of the gels and were only assigned when a spot was clearly present in all gels of the group. After gel spot matching, a scatter plot was created between pairs of gels from each subgroup of replicates for a particular loading method, and a correlation coefficient was calculated (see discussion section). More detailed comparisons of spot quantities within and between the different sample loading methods, were carried out using only the landmarked spots. The reproducibility of spot quantity for each landmarked spot was calculated across the replicates after exporting the un-normalized values to an Excel spreadsheet (Microsoft, Redmond, WA, USA). The mean quantity values and overall average coefficients of variation of spot quantity are reported (CV, standard deviation divided by the mean). It should be noted that by averaging CV values, the range in variability for each individual spot quantity is not shown, but by using 30-100 CV values for each loading method we can approach the true variability of the method.

Results

Comparison of the Number of Spots Detected

The reproducibility of each sample loading method was investigated by measuring the number of spots detected in 2-DE separations on IPG 6-11's and IPG 4-7's. Representative 2-D gel images for each of the sample loading techniques are displayed in

Figure 19 for groups A-D. The number of spots detected in each gel was tabulated and the mean, standard deviation, and coefficient of variation (%CV) calculated for each set of replicates (Table 4). For IPG 6-11's, cup-loading experiments resulted in the greatest number of spots detected with the lowest overall variations. The number of spots detected by cup loading ranged from 772 to 935 over 11 gels, with a mean value of 865. The %CV ranged from 2-8%, with an average variation of 5.2%. Active rehydration of the same sample extracts resulted in 71% as many spots, on average, relative to the mean number from cup-loading. A wide range of 348 to 1108 spots resulted in a mean value of 617 and a %CV range of 12-22%. The greater variability in spot numbers for actively rehydrated gels appears to be a direct result of greater horizontal and vertical streaking.

Both passive rehydration methods resulted in lower CV's than active rehydration and higher CV's than cup-loading. Passive rehydration using the in-tube method had a CV range of 6-9% while the in-tray method was more variable at 11-18%. However, in-tube passive rehydration (Figure 21, A and C) resulted in 20% fewer detected spots than the passive in-tray method (Figure 21, B and D) and 32% fewer than cup-loading. Surprisingly, in-tube passive rehydration resulted in roughly the same average number of spots (591) as active rehydration (617), which was approximately 2/3 as many as detected in cup-loading experiments. On the other hand, the in-tray passive method resulted in 17% more spots on average than the in-tube or active rehydration separations. The average number of spots detected for the in-tray passive rehydration method (726) was 84% of the cup-loading numbers (865).

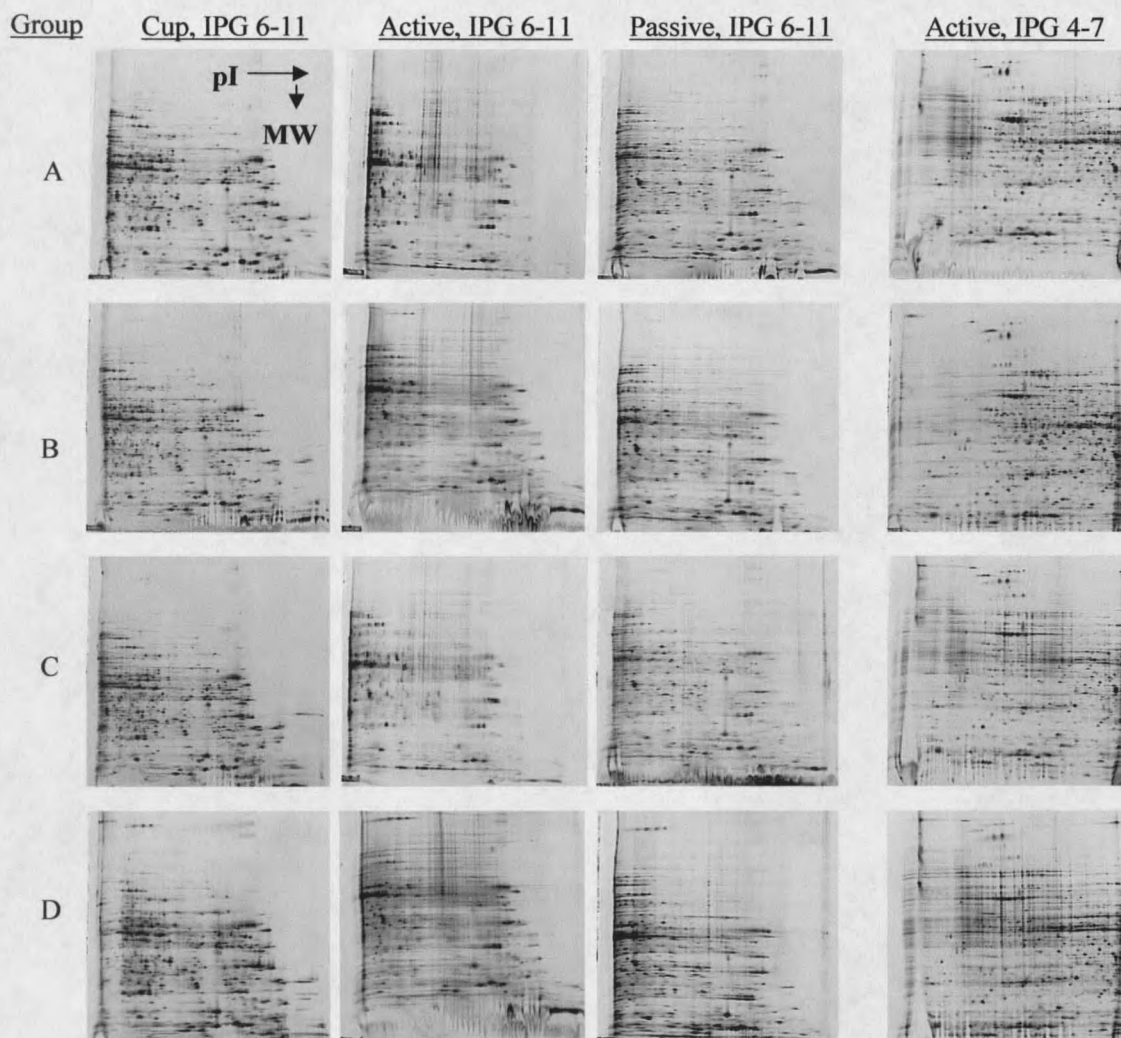


Figure 21. Representative 2-D gel images of samples A-D, each separately prepared and resolved over IPG 6-11 and IPG 4-7 ranges. Sample was applied to the IPG strip using either cup-loading, active rehydration, or passive rehydration. Passive rehydration for samples A and C were performed in 2mL pipet tubes and for samples B and D in the same tray as used for IEF. SDS-PAGE performed using 8-18% gradient gels. Gels were stained with SYPRO Ruby and imaged using the Molecular Imager FX (Bio-Rad).

Table 4. Number of spots detected using different sample loading methods and PDQuest^{a)}

IPG 6-11					IPG 4-7					
Group		Cup	Active	Passive	Group		Active	Group		Active
				2mL tubes						
A ^{b)}	Mean	772	641	465	A ^{b)}	Mean	760	E ^{b)}	Mean	812
	SD	64	131	27		SD	36		SD	118
	%CV	8.3	20.4	5.8		%CV	4.7		%CV	14.6
				IEF tray						
B ^{b)}	Mean	850	533	565	B ^{b)}	Mean	1124	F ^{b)}	Mean	1020
	SD	50	104	101		SD	102		SD	69
	%CV	5.9	19.5	17.8		%CV	9.1		%CV	6.8
				2mL tubes						
C ^{b)}	Mean	924	348	716	C ^{b)}	Mean	1067	G ^{b)}	Mean	920
	SD	42	78	65		SD	46		SD	71
	%CV	4.5	22.4	9.1		%CV	4.3		%CV	7.7
				IEF tray						
D ^{c)}	Mean	935	1108	887	D ^{b)}	Mean	1334	H ^{b)}	Mean	1119
	SD	18	128	101		SD	82		SD	107
	%CV	2	11.5	11.4		%CV	6.2		%CV	9.6
A-D				tubes/tray			A-D	E-H	A-H	
Mean		865	617	591/726		Mean	1071	968	1020	
a) The same spot detection parameters were used across all gels in a group.										
b) Spot detection and values were calculated from three replicate gel images.										
c) Spot detection and values were calculated from duplicate gel images.										

The same independently isolated samples (A-D) were also focused on IPG 4-7's, using the active rehydration method, in order to compare the amount of variation between separating the acidic proteins of a mixture with separating the basic proteins from the same mixture. Further, an additional set (E-H) of independently isolated samples was run to further assess possible variability caused by sample preparation. The relative dispersion of spot numbers detected for samples A-D ranged from 4-9%, E-H 7-15%, and

overall, A-H, 4-15% for the eight sets of triplicates from eight separate preparations (Table 4).

Examination of all of the individual gels of the groups A-H suggest that the range in CV values in numbers of spots detected (4-15%) resulted mostly from falsely assigning spots to streaks on the acidic side of the gels. This type of streaking was variable for different individually prepared samples. It seems plausible that different amounts of nonprotein impurities (such as lipids) may be causing the streaking, which may be a consequence of using sonication during cell lysis. As compared to their basic counterparts, the range in CV's of spots detected for acidic separations was slightly more than for cup-loading (2-8%) and in-tube passive rehydration (6-9%), and lower than the in-tray passive rehydration (11-18%) or active rehydration methods (12-22%). It should be noted that generally more proteins were resolved over the same 17.5 cm distance when using IPG 4-7's as IPG 6-11's; and yet, given this greater complexity, the reproducibility of the acidic separations was more similar to the basic separations using the cup-loading technique than using the same active rehydration protocol with IPG 6-11 strips. Clearly, the coefficient of variation in spot numbers for active rehydration of the more acidic proteins is consistently lower than active rehydration for the basic counterpart, and approaches a similar reproducibility and the high spot number counts as cup-loading with IPG 6-11's.

Correlation Coefficients, Spot Matching, and Matching Efficiency

To further assess reproducibility, pairs of gels from a replicate set representing a specific loading method were matched against each other in order to determine the number of spots focused to identical locations (x,y) and to calculate the correlation coefficient, r , of the quantities from those identical spots for that pair of gels (Table 5). Scatter diagrams were created using the Scatter Plot function of PDQuest after spot detection and the assignment of landmarks. This function plots the quantity of each spot from one gel against the quantity of the matching spot in a second gel on a log scale. If all of the matched spots have the exact same quantity, the spots will be plotted on a line with slope 1.0. Further, matched spots between a pair of gels whose quantities differ will result in correlation coefficients less than 1.0. The Scatter Plot function of PDQuest outputs two values: the number of spots found matching between the two gels and the correlation coefficient among those spots. Therefore, calculating the correlation coefficient and determining the number of spots matched between any two gels is a good way of determining how similar the quantities are for the matched spots and how many proteins in each gel were reproducibly focused to the same identifiable location.

Table 5. Correlation coefficients between replicate gels for each sample loading method

IPG 6-11 ^{a)}							IPG 4-7 ^{b)}					
Group	Cup	N ^{c)}	Active	N	Passive	N	Group	Active	N	Group	Active	N
				2mL tubes								
A	0.726	647	0.659	96	0.777	187	A	0.839	384	E	0.739	354
	0.714	454	0.765	90	0.68	114		0.749	459		0.809	394
	0.739	380	0.769	72	0.61	118		0.752	451		0.786	319
				IEF tray								
B	0.819	574	0.397	83	0.747	147	B	0.709	855	F	0.69	617
	0.774	498	0.53	106	0.778	131		0.647	666		0.753	684
	0.795	417	0.63	84	0.822	154		0.731	762		0.622	554
				2mL tubes								
C	0.641	469	0.674	44	0.786	117	C	0.636	625	G	0.801	530
	0.566	327	0.679	43	0.838	134		0.689	505		0.745	498
	0.526	511	0.687	44	0.748	117		0.772	772		0.833	474
				IEF tray								
D ^{d)}	0.676	577	0.56	121	0.595	106	D	0.672	889	H	0.746	622
					0.653	104		0.65	895		0.681	606
					0.63	146		0.563	622		0.654	536
A-D							A-D			E-H		
Mean	0.698	485	0.635	78	0.737	131	Mean	0.701	657	Mean	0.738	516
				Tray 0.70								
Detected ^{e)}		865		617	Tube	591	A-H					
					Tray	726	Mean	0.722	586			
Matching ^{f)}						Detected ^{e)}			1020			
Efficiency		56%		13%	Tube	22%			57%			
					Tray	18%						

a) Scatter plots were created and correlation coefficient, r , calculated after establishing 30-40 landmarks.

b) Scatter plots were created and r calculated after establishing 55-100 landmarks.

c) N represents the total number of matched spots between two replicates.

d) Only duplicates used in cup-loading and active rehydration of D, IPG 6-11, triplicates in all others.

e) Average number of spots detected, from Table 1.

f) Matching efficiency is defined as the mean number of spots matched divided by mean number detected.

In order to properly distinguish differences between the loading methods, we had to take into account the possible variations that exist between the different groups of sample preparations and among the groups (since each sample was individually prepared and multiple sample handling steps are involved, respectively). To assess these variations, statistical mixed models were used. Mixed models allow one to assign fixed variables and random variables in a data set to determine the variance (standard deviation squared) that these variables contribute to the experimental result (Pinheiro et al, 2000). In our study, the fixed variables are the three different sample-loading methods: cup-loading, active and passive rehydration. We partitioned random variables into two groups: 1) to subtle differences that may exist between sample preparations, and 2) to an "all other" variable that may cause variance among the replicates for a given method. For example, if the amount of membrane fragments in a sample varied between sample preparations, because of differences in sonication time and power, one may find that the results across one group of samples is different than across another group of samples. Alternatively, differences in the amount of protein loaded between the replicates could also conceivably influence the final results. Thus, by using statistical mixed models to individually evaluate the correlation coefficients and the numbers of spots matched between gels, we can thoroughly test for differences between sample loading methods after taking into account any variations between the sample preps and among the replicates.

Across all of the IPG 6-11 gels, the estimated random effect on the correlation coefficient values as a result of using different sample preparations (A-D) was estimated

to be a variance of 2.0×10^{-5} with a corresponding 95% confidence interval of (0.0, 8.97×10^{31}). This low value indicates that there was very little variance that could be attributed to sample preparation. However, the corresponding 95% CI is extremely wide and suggests that our information is very limited from using only four preparations. Thus, we evaluated the effect of sample variation more thoroughly by running eight different sample preparations on IPG 4-7's, as discussed below. Variability attributed to "all other" random effects when producing the replicates for a method (using IPG 6-11's) was estimated at 9.7×10^{-3} , with a 95% CI of (0.00578, 0.0167). This indicates that variance between replicates influenced the overall correlation coefficients approximately 500 times more than variance between sample preparations. By incorporating the random estimates and using the null hypothesis to evaluate the three loading methods, an F-value of 1.60 and $P=0.2147$ was generated. The high p-value indicates that there is no significant difference between mean correlation coefficient values obtained for the three loading methods. A visual inspection of the average correlation coefficients for quantitation for each method shows that they are all about 0.700 (Table 5).

On the other hand, comparisons between the numbers of spots matched between any pair of gels from a set of replicates (used in calculating the correlation coefficients of the IPG 6-11 gels) was directly attributable to the loading method used. Random variance between sample preparations was estimated at 274.9 with a 95% CI of (2.056, 36760.7) and random variance between replicate samples was 3319.7, with a 95% CI of (1901.5, 5795.7). An F-value of 100.35 and $P < 0.0001$ was calculated providing statistical evidence that the number of spots typically found matching was directly related

to the sample method used. From Table 5, gels obtained from cup-loading averaged 485 matching spots as compared to 131 for passive rehydration and 78 for active rehydration.

Building upon the significance of the number of spots typically matched between replicate gels for a particular loading method, we define the term "matching efficiency" as the average number of spots found matching between a pair of replicate gels (for a particular loading method) divided by the average number of spots detected across the replicate gels (for that method) (Table 5). The matching efficiency generated by the image analysis software is useful to assess gel-to-gel reproducibility. For the gels obtained by in-tube passive rehydration, only 22% of the spots detected on these gels were matched between replicates (an average 131 spots matched out of an average total spot count of 591). The in-tray passive rehydration method had a matching efficiency of 18% (131 average spots matched out of an average 726 detected). Results for replicates obtained from active rehydration were poor with only 13% of the spots being matched and scatter plotted (an average of 78 matched out of an average 617 detected). By contrast, cup-loading replicates matched dramatically better at 56% (an average of 485 spots out of an average 865 detected). When the sample extracts were separated over IPG 4-7, using active rehydration, an almost identical matching efficiency, 57%, was obtained (Table 5). Thus, the gel-to-gel reproducibility of cup-loading IPG 6-11 experiments as judged by matching efficiency are very similar those obtained by active rehydration over IPG 4-7.

Affects of Different Sample Preparations and Repeat Gels on Matching Efficiency

The influence of random variables on correlation coefficients and matching numbers was further evaluated by running sets of triplicates from eight different sample preparations over IPG 4-7's using active rehydration. For random effects on correlation of quantity, prep-to-prep variance was estimated at 0.00242 (95% CI (0.00553, 0.0106)) and the variance between replicates was 0.00139 (95% CI (0.00279, 0.00558)). Therefore, the total random variance was 0.00381 and roughly 2/3 of it was attributed to differences obtained during sample preparation (0.00242/0.00381). The standard deviation of the correlation coefficients within a single prep was 0.0373. These numbers indicate that correlation coefficients, which reflect spot quantity reproducibility, can vary by roughly +/- 5% between sample preparations (0.0373 divided by an average correlation coefficient of 0.700) and by roughly +/- 3.5% between replicates (0.049/0.700). Here again, the influence of random variances on correlation coefficients is rather small.

On the other hand, the numbers of spots that are typically matched between a pair of replicate gels and, hence, gel-matching efficiencies, were directly affected by the two types of random variables. Within a set of replicates, random variance from non-sample prep variables resulted in a standard deviation of 88 matching spots. This means the matching efficiency of IPG 4-7 gels could range from 49-66% (average number of spots matched, 586, +/- 88 divided by the average number of spots detected, 1020). Compared to the 57% matching efficiency given earlier, this suggests that matching efficiency could

vary by 9% just from repeating the experiments while using the same sample extract. The impact of separate sample preparations on variance was three times larger than the non-sample prep variance. A sample prep standard deviation of 145 equates to a matching efficiency that can range from 43-72%. This means variations due to sample preparations could affect matching efficiency numbers by as much as +/-15%.

Reproducibility of Spot Quantity within a Sample Loading Method

For each group of gels (A-H), thirty or more landmarked spots were selected and the quantity of each of these spots was compared between replicate gels and between the different sample loading methods. The same image analysis parameters were used across all gels within a particular group. The quantity for each landmarked spot was calculated by Gaussian fitting during spot detection, using the PDQuest Spot Detection Wizard. PDQuest uses the following formula in calculating spot quantity: $\text{Spot height} * \pi * \sigma_x * \sigma_y$, where spot height is the maximum value of the Gaussian fit to the spot, σ_x is the standard deviation of the spot (width) along the x-axis, and σ_y the standard deviation of the spot (width) along the y-axis. The quantities of the landmarked spots were exported to an Excel spreadsheet and the mean value, standard deviation, and coefficient of variation computed. An average CV was tabulated from all of the landmarked spots for each loading method and is shown in Table 6. To compare the overall average CV values between the methods, a comparison of means test was used. Finally, by comparing the

actual quantity of the spot between the three different loading methods, we determined if any bias exists for a particular method to focus specific proteins.

Table 6 shows the average coefficient of variation for each of the loading methods per sample group. For IPG 6-11 separations, cup-loading experiments produced the lowest average spot quantity CV between replicate gels at 27.7% while active rehydration methods produced the highest at 41.4%. Spot quantity CV's for the passive methods were intermediate to those of cup-loading and active rehydration at 32.3% for the in-tube method and 37.0% for the in-tray method. Acidic separations, on IPG 4-7's, were similar to cup-loading results with an average quantity CV of 23.8% (A-H). An ANOVA procedural test was used to determine if there was an overall difference in the means, collectively. An F-value of 6.600 and P-value of 0.034 was calculated and shows that the mean CV's of the four methods are not all equal. To determine where those differences occur, multiple comparisons were made using Tukey's method (Neter et al, 1996). Tukey's method uses the estimate of variance pooled across all groups and avoids situations of using $n = 2$, which would have occurred when comparing the in-tube and in-tray passive rehydration means to the other means. Further, it protects against the problem of non-conservative p-values, due to multiple simultaneous tests. At the 5% significance level (95% confidence interval) the only differences found were between IPG 6-11 cup-loading/IPG 6-11 active rehydration and between IPG 6-11 active rehydration/IPG 4-7 active rehydration. By inspection of the means in Table 6, the average coefficient of variation for IPG 6-11 active rehydration was 1.5 times larger than IPG 6-11 cup-loading and 1.7 times larger than IPG 4-7 active rehydration.

Table 6. Average coefficient of variation for landmarked spot quantities within replicate gels.

IPG 6-11 ^{a)}									IPG 4-7 ^{a)}		
Group	Cup	N ^{b)}	Active	N	Passive	N	Passive	N	Group	Active	N
				2mL tubes			IEF tray				
A	0.265	35	0.442	35	0.271	35	n.a.		A	0.225	100
B	0.267	40	0.451	40	n.a. ^{d)}		0.355	40	B	0.269	55
C	0.298	30	0.379	30	0.375	30	n.a.		C	0.23	82
D	0.28	36	0.384	36	n.a.		0.384	36	D	0.254	62
									E	0.244	100
									F	0.262	55
									G	0.193	82
									H	0.225	62
									A-D	E-H	A-H
Mean	0.277		0.414		0.323		0.37		0.244	0.231	0.238
95% CI ^{e)}	0.015		0.037		0.102		0.029		0.02	0.029	0.017

a) Coefficient of variation is not shown for individual spot quantities due to space limitations.

Instead, a mean CV value was calculated for the N quantities in order to represent the average dispersion for a spot quantity when using a particular loading method.

b) N is the number of landmarked spots selected for quantitation and subsequent matching.

c) Only IPG 4-7 triplicates run for E-H.

d) Specific type of sample loading not applicable (n.a.) for this group.

e) Only the mean values were used in calculating the 95% CI.

In all cases (both IPG 6-11 and IPG 4-7) the range in variance for individual spot quantities, generally ranged from 3-90% (for the landmarked spots). Both the IPG 6-11 cup-loading and IPG 4-7 active rehydration experiments yielded larger protein quantities, which correlated with lower CVs (see next section). Surprisingly, landmarked spots which had a large CV value for one loading method often had a much lower CV value in the other loading methods, which suggests that each loading technique may have a bias for specific proteins. Overall, the wide range in CV's for spot quantity provides clear

evidence of the need for statistical analysis when comparing protein expression differences between 2-D gels. Algorithms that can calculate p-values (e.g. < 0.05) for spot quantities would be preferred over simple 2-fold or other ad hoc cut-off values for evaluating expression differences in proteomic studies.

Spot Quantity Comparisons between Different Sample Loading Techniques

To investigate the possible selectivity of a method for loading certain types of proteins over an IPG 6-11 range, the quantity of each landmarked spot was compared, side-by-side, between the three loading methods (attached as an addendum below). The average quantity for a spot was listed for each loading method and the relative quantities between methods was calculated (Table 7A). In general, average quantities obtained by active rehydration were 74% of those from cup-loading. Average quantities obtained by passive in-tube rehydration showed the lowest abundance, yielding only 47% the amount obtained from cup-loading and essentially the same amount, 94%, as seen with active rehydration. Surprisingly, in-tray passive rehydration registered 35% more protein on average than cup-loading and 129% more than active rehydration. A closer inspection of the quantities across the experiments demonstrates that certain proteins were recovered and focused better with one technique than another (Table 7B).

Table 7. Comparison of protein quantities in landmarked spots from samples A-D, IPG 6-11's

A.					
Group	N ^{a)}	Quant. Act./ Quant. Cup ^{b)}	Quant. Pass./ Quant. Cup ^{b)}	Quant. PasQuant. Pass./ Quant. Active ^{b)}	
A	35	58	34	85	
B	40	78	136	225	
C	30	79	59	103	
D	35	82	134	233	
Mean		74	47 ^{c)} 135 ^{d)}	94 ^{e)} 229 ^{d)}	
B.					
Group	N ^{a)}	No. of Active>Cup ^{e)}	No. of Pass. >Cup ^{e)}	No. of Pass. & Act. > Cup ^{e)}	No. of Pass. >Active ^{e)}
A	35	5	1	1	7
B	40	11	19	9	30
C	30	6	6	2	9
D	35	11	22	10	24
Totals	140	33 ^{f)} 75 ^{g)}	7 ^{f)} 41 ^{g)}	3 ^{f)} 19 ^{g)}	16 ^{f)} 54 ^{g)}
a) Number of landmarked spots selected for quantitation.					
b) Average over N landmarks for the ratio of spot quantity from one method relative to its value in another method. Abbreviations: Quantity (Quant.), Active rehydration (Act.), Cup-loading (Cup), Passive rehydration (Pass.). Percent values are given.					
c) Passive rehydration using 2mL tubes, groups A and C only.					
d) Passive rehydration using the in-tray method, groups B and D only.					
e) The number of landmarked spots which were higher in abundance using one method versus another.					
f) Totals are from groups A and C.					
g) Totals are from groups B and D.					

An average of 23% (33/140) of the landmarked proteins had higher quantities when active rehydration was used as compared to cup-loading. Of these 23%, 4 of 5 in exp. A and 4 of 6 in exp. C also had higher values than the corresponding passive in-tube

method and 5 of 11 in exp. B and 4 of 11 in exp. D were higher than the corresponding in-tray values. Fourteen of these seventeen proteins (which achieved their highest abundance from using active rehydration) were found to have pI's > 8 and MW's $< 50,000$ Da, while three represented high molecular weight proteins (in exp. B only), one approximately 200,000 Da and two at 80,000 Da.

In-tube passive rehydration displayed one protein out of 35 in experiment A with higher abundance than cup-loading and 6 of 30 in experiment C. Of these seven, only three were similarly in higher abundance with active rehydration. Most likely, these three proteins tend to precipitate more readily in concentrated solutions, such as those that develop during cup-loading. Overall, the lowest protein quantities typically occurred using the in-tube passive rehydration method and is likely due to protein adsorption to the walls of the plastic tubes.

In experiments where samples were passively rehydrated in the same IEF tray as used for focusing (experiments B and D), 55% (41 of 75) of the landmarked spots registered greater abundance than cup-loading and 72% (54 out of 75) greater than active rehydration. Thus, passively rehydrating sample using the in-tray method over IPG 6-11 focused 84% as many proteins as cup-loading and 55% of those proteins were in greater abundance as compared to the cup method. However, there was also 10% more spot quantity variance with the in-tray method as compared to the cup method. Recall that this in-tray passive rehydration method included 24 hrs of rehydration with no voltage followed by a voltage program to 87 KVh, where the first two steps included 100 V for 2 h and 300 V for 6 h. It may be that this method acts as a hybrid between purely passive

rehydration and active rehydration and consequently any proteins poorly absorbed passively may be better absorbed electrophoretically. From the data, half of the proteins in higher abundance using the in-tray method were also higher during active rehydration, which suggests that focusing these proteins is either enhanced by the application of a low voltage or that these proteins have a tendency to precipitate under concentrated conditions, such as those which develop during cup-loading. Since streaking appears to be the least in the cup-loading experiments, it is more likely that the applied voltage helped in loading those proteins.

Discussion

The quantitative and statistical data presented demonstrates that the best way to separate the more basic components of a protein mixture over IPG 6-11 is with cup-loading. This is in spite of the potential problem of protein precipitation at the sample application point, especially when a large amount of protein is loaded. A homemade set-up was used in this study to cup-load 300 μg of protein in 125 μL volume with minimal precipitation and vertical streaking. This technique has now been successfully used in our lab over 60 times and only requires careful placement of the cup onto the rehydrated IPG strip. For IPG 6-11, cup-loading provided the greatest number of spots (865) with the lowest variances (2-8%), most reproducible spot quantities among replicate gels (28% CV), and by far the highest matching efficiencies between replica gels, at 56% (calculated as the number of spots matched divided by the average number of spots detected). While

this matching efficiency may seem low, we could not find other studies that defined matching efficiency as used here and in those studies where the term was used it was unclear exactly how the term was defined and so comparisons were difficult to make. It seems likely, however, that lower protein loading could lead to higher matching efficiency, but we did not investigate this since we were interested in larger protein loads to facilitate subsequent mass spectral analysis of lower abundance proteins. It was also determined that gel matching efficiency could vary by as much as 15% between sample batches as a likely consequence of using a sonication lysis procedure.

Active rehydration results were surprisingly poor in comparison to the other methods. While 71% as many spots were detected as compared to cup-loading, replicate gels produced the poorest matching with only 13% of the spots detected being matched between replicates. Further, greater spot quantity variances were seen with this method than with any other, at 41%. This method, however, did seem to have some favorable bias toward more basic proteins with moderate molecular weights. Seventeen landmarked proteins out of a total of 140 had their highest abundance when active rehydration was used. Of these, 14 had $pI > 8$ and $MW < 50\text{KDa}$, and three had $pI < 8$ with two having MW 's of $\sim 80\text{KDa}$ and one of $\sim 200\text{KDa}$. These particular high MW proteins were found on all the gels from sample groups A-D, but only showed higher abundance in sample group B.

It is not entirely clear why sample loading with active rehydration produced such poor IPG 6-11 gels. In all cases, the same concentration of carrier ampholytes (0.4%), detergents, and thiourea (1.5 M) were used to minimize protein losses and improve

protein solubility and resolution. Practical experience indicates that more extensive sonication during cell lysis produces more of this streaking, which is consistent with the presence of more membrane fragments and lipids. For this and future studies, sample prefractionation was intentionally avoided in order to minimize the total number of gels needed for differential gel analyses. Therefore, ultracentrifugation was not used to remove high-molecular weight fragments, and it might be that this streaking effect was due to the presence of contaminants such as lipids.

Two different types of passive rehydration methods were evaluated. One in which the IPG strip was rehydrated with buffer containing sample within snug-fitting tubes cut from 2 mL serological pipettes and the other within the same IEF tray as used for focusing. The in-tube method out-performed active rehydration and the in-tray method in terms of spot number variability (6-9%), gel-to-gel correlation (0.737), and average variations in spot quantity (32%). However, in-tube passive rehydration provided the fewest number of spots (591) and lowest spot abundances (47% as much as cup-loading) than any of the other methods. Losses during passive rehydration was studied by Zuo and Speicher who found that 5% of radiolabeled proteins were lost by adsorption to the plastic walls of the rehydration chamber when loading 50 μg of proteins (Zuo and Speicher, 2000). They also found that larger protein loads led to larger losses and that as much as 50% protein loss could occur during passive rehydration when loading 500 μg of protein (they also used 450 μL total rehydration buffer and excess liquid remained unabsorbed). Presumably as much as 30% of the 300 μg protein loading applied in this study could be lost to the chamber during our in-tube passive rehydration, but the losses

would necessarily have been consistent enough to permit the reproducibility results seen in this study.

The in-tray method, on the other hand focused 84% as many spots as cup-loading (%CV of 11-18%) with 55% of these proteins being of higher abundance than cup-loading. It seems that with the in-tray method, proteins that do not absorb well into the strip may be better enticed into the strip with the addition of low voltage at the beginning of the IEF run. Then again, spot quantities varied on average by 37%, which was 10% than the cup-loading results. It is unlikely that spot quantity variations were due to absorption losses to the wicks or chamber of the IEF tray since active rehydration over IPG 4-7 produced similar spot quantity variations (24%) to IPG 6-11 cup-loading (28%), unless basic proteins are preferentially absorbed to the IEF tray more so than acidic ones.

For each of the methods used in this study, some of the variability witnessed in the spot quantity values might be related to how the PDQuest software calculates protein abundance. PDQuest creates a Gaussian representation from the pixel intensities and subtle variations in spot contours and boundaries may be reflected as variations in spot quantity. It is possible that lower spot quantity variations could have been achieved had each landmarked spot boundary been individually optimized on every gel within every group, or had a higher resolution been used during image acquisition (e.g. 50 μM instead of 200 μM). However, any systematic error(s) in calculating spot quantities would presumably have been the same across all gels within a group and, thus, it seems likely that subsequent comparative analyses would have yielded the same overall trends as reported for each of the loading methods.

Other types of software, such as Progenesis, use more of a topographical algorithm for spot measurement which has been reported to be more precise (Nishihara and Champion, 2002). However, in a comparison between Progenesis, PDQuest and Z3, Nishihara and Champion recently evaluated the reproducibility of quantitation for twenty selected spots from four replicate gels (each loading 30 μg of protein) and reported the range in variability (%CV) to be essentially the same among all three programs, at 3-33% (Nishihara and Champion, 2002). In this study, the range in variability in spot quantity in all cases was generally between 3-90% when loading 300 μg of protein. This rather wide range in variability, with both high and low protein loading, indicates that rigorous statistical methods should be used when analyzing protein expression differences. It was also reported that SYPRO Ruby staining was not equivalent for all proteins by demonstrating different limits of detection for certain types of proteins (Nishihara and Champion, 2002). We rule out the possibility that differences in fluorescent staining is the cause of variability in quantitation because the same proteins were landmarked and matched to all gels within an experiment group, and so one can expect the identical proteins on each gel to stain in an identical fashion (when using the same stock solutions).

Concluding Remarks

Quantitative statistical analyses supporting one sample loading method for 2-DE over another have been limited. In particular, past quantitative analyses have been primarily applied to broad-range separations (pH 3-10) using radiometric methods. With

the development of new IPG strips for higher resolution over shorter pH ranges, investigators can now load larger amounts of protein in order to view more of the proteome over a given pH range. Having a firm idea of how to best load those proteins will save resources and provide better results. In this study, we quantitatively evaluated three different sample-loading techniques for resolving basic proteins over IPG 6-11. We used SYPRO Ruby and PDQuest to compare the spot numbers, spot reproducibility and spot quantities of cup-loading to those of active and passive rehydration methods and compared these results with the reproducibility of separating the acidic components of the same mixtures over IPG 4-7, using active rehydration. The cup-loading sample application method greatly out-performed active and passive rehydration methods for separation of basic proteins by consistently providing the greatest number of detectable spots, best gel-to-gel reproducibility, lowest spot quantity variations, and, qualitatively, the best looking gels. Most importantly, replicas from cup-loading gels were far superior to the other methods in providing the best gel matching efficiencies. The data for cup loading also closely matched the high spot numbers, matching efficiency, and reproducibility obtained for acidic separations over IPG 4-7, using active rehydration. Our analysis of spot quantity variation clearly demonstrates the need for statistical rigor when comparing protein expression differences between 2-D gels. If cut-off levels are to be used, p-values (e.g. < 0.05) will be much more reliable than 2-fold or other ad hoc cut off values for studying expression differences in proteomic experiments.

Addendum

Table 9. Quantity values for landmarked spots from sample B, IPG 6-11's^{a)}

Landmark	In-Tray Method					
	Cup Quant.	Active Quant.	Passive Quant.	Active/Cup	Passive/Cup	Passive/Active
1	2.80E+06	2.14E+06	9.29E+06	0.76	3.32	4.35
2	6.35E+06	7.52E+06	1.08E+07	1.18	1.7	1.43
3	3.51E+06	1.54E+06	2.00E+06	0.44	0.57	1.3
4	4.22E+06	2.40E+06	4.05E+06	0.57	0.96	1.69
5	1.14E+07	8.59E+06	1.97E+07	0.75	1.72	2.29
6	3.25E+06	2.95E+06	4.74E+06	0.91	1.46	1.61
7	9.00E+06	4.83E+06	4.31E+06	0.54	0.48	0.89
8	4.10E+06	2.47E+06	6.20E+06	0.6	1.51	2.51
9	5.44E+06	7.00E+06	1.02E+07	1.29	1.88	1.46
10	5.57E+06	4.67E+06	4.36E+06	0.84	0.78	0.93
11	2.50E+06	2.83E+06	2.23E+06	1.13	0.89	0.79
12	8.44E+05	2.39E+05	2.89E+06	0.28	3.42	12.09
13	5.06E+06	2.23E+06	1.67E+06	0.44	0.33	0.75
14	5.46E+06	3.30E+06	6.63E+06	0.6	1.21	2.01
15	3.85E+06	2.30E+06	6.51E+06	0.6	1.69	2.83
16	4.30E+06	4.34E+06	4.01E+06	1.01	0.93	0.92
17	3.52E+06	2.80E+06	6.75E+06	0.79	1.92	2.41
18	5.57E+06	5.10E+06	1.43E+07	0.92	2.57	2.8
19	9.99E+06	5.45E+06	2.18E+07	0.55	2.18	4
20	6.58E+06	2.97E+06	6.41E+06	0.45	0.97	2.16
21	6.11E+06	6.02E+05	4.63E+06	0.1	0.76	7.69
22	4.44E+06	1.63E+06	4.31E+06	0.37	0.97	2.65
23	5.23E+06	5.45E+06	1.78E+07	1.04	3.4	3.26
24	1.03E+07	1.13E+07	1.05E+07	1.1	1.03	0.94
25	6.60E+06	1.51E+07	2.35E+07	2.29	3.56	1.55
26	4.82E+06	5.51E+05	3.96E+06	0.11	0.82	7.18
27	1.83E+06	1.62E+06	2.45E+06	0.88	1.34	1.51
28	9.22E+06	3.80E+06	9.01E+06	0.41	0.98	2.37
29	6.26E+06	1.93E+06	2.26E+06	0.31	0.36	1.17
30	7.23E+06	2.46E+06	4.33E+06	0.34	0.6	1.76
31	1.73E+07	2.35E+07	2.55E+07	1.36	1.47	1.08
32	4.53E+06	3.92E+06	3.96E+06	0.87	0.87	1.01
33	3.56E+06	1.71E+06	2.09E+06	0.48	0.59	1.23
34	3.80E+06	6.58E+05	5.77E+05	0.17	0.15	0.88
35	2.90E+06	1.14E+06	1.34E+06	0.39	0.46	1.18
36	6.88E+06	8.93E+06	1.85E+07	1.3	2.69	2.07
37	4.46E+06	3.94E+06	4.34E+06	0.88	0.97	1.1
38	4.63E+06	3.85E+06	2.78E+06	0.83	0.6	0.72
39	2.94E+06	6.17E+06	3.40E+06	2.1	1.16	0.55
40	5.23E+06	6.75E+06	6.01E+06	1.29	1.15	0.89
			Mean	0.78	1.36	2.25

a) Quantity values are average values. The ratios are calculated between methods.

Table 8. Quantity values for landmarked spots from sample A, IPG 6-11's^{a)}

Landmark	2mL Tubes					
	Cup Quant.	Active Quant.	Passive Quant.	Active/Cup	Passive/Cup	Passive/Active
1	9.36E+06	4.48E+06	2.83E+06	0.48	0.3	0.63
2	1.27E+07	1.20E+07	5.32E+06	0.95	0.42	0.44
3	5.88E+06	2.51E+06	9.70E+05	0.43	0.16	0.39
4	7.81E+06	2.77E+06	5.21E+05	0.35	0.07	0.19
5	2.24E+07	1.01E+07	8.32E+06	0.45	0.37	0.83
6	3.54E+06	4.59E+06	1.12E+06	1.3	0.32	0.24
7	9.52E+06	5.67E+06	7.57E+05	0.6	0.08	0.13
8	1.40E+06	3.62E+06	2.46E+06	2.59	1.76	0.68
9	7.59E+06	4.07E+06	1.32E+06	0.54	0.17	0.32
10	1.23E+07	3.82E+06	4.54E+06	0.31	0.37	1.19
11	8.23E+06	3.33E+06	2.83E+06	0.4	0.34	0.85
12	2.99E+07	3.30E+06	5.73E+06	0.11	0.19	1.74
13	2.82E+06	8.36E+05	8.04E+05	0.3	0.28	0.96
14	8.94E+06	4.79E+06	3.38E+06	0.54	0.38	0.71
15	1.31E+07	2.17E+06	2.72E+06	0.17	0.21	1.25
16	5.88E+06	1.90E+06	7.82E+05	0.32	0.13	0.41
17	3.65E+06	1.49E+06	1.18E+06	0.41	0.32	0.79
18	2.23E+07	6.37E+06	5.38E+06	0.29	0.24	0.84
19	3.14E+07	1.03E+07	9.59E+06	0.33	0.31	0.93
20	2.97E+07	7.91E+06	6.37E+06	0.27	0.21	0.81
21	1.05E+07	9.96E+05	1.96E+06	0.09	0.19	1.97
22	9.81E+06	2.33E+06	1.74E+06	0.24	0.18	0.75
23	1.42E+07	1.22E+06	4.93E+06	0.09	0.35	4.06
24	1.14E+07	2.68E+06	2.28E+06	0.23	0.2	0.85
25	2.70E+07	6.06E+06	5.86E+06	0.22	0.22	0.97
26	1.85E+06	4.32E+06	1.53E+06	2.33	0.83	0.35
27	4.49E+06	4.42E+06	3.80E+06	0.99	0.85	0.86
28	4.44E+06	5.70E+06	1.99E+06	1.28	0.45	0.35
29	5.06E+06	6.78E+06	3.79E+06	1.34	0.75	0.56
30	6.28E+06	3.74E+06	2.49E+06	0.6	0.4	0.67
31	6.92E+06	2.26E+06	6.88E+05	0.33	0.1	0.31
32	5.59E+06	3.49E+06	9.90E+05	0.62	0.18	0.28
33	7.72E+06	1.35E+06	1.98E+06	0.18	0.26	1.46
34	2.38E+07	8.82E+06	2.33E+06	0.37	0.1	0.26
35	8.74E+06	2.10E+06	3.45E+06	0.24	0.39	1.64
			Mean	0.58	0.34	0.85

a) Quantity values are average values for each landmarked spot. The ratio of these values between two different sample loading methods is also calculated.

Table 10. Quantity values for landmarked spots from sample C, IPG 6-11's^{a)}

Landmark	Cup Quant.	Active Quant.	2mL Tubes		Passive/Cup	Passive/Active
			Passive Quant.	Active/Cup		
1	5.21E+06	9.15E+06	9.23E+06	1.76	1.77	1.01
2	3.54E+06	2.12E+06	8.74E+05	0.6	0.25	0.41
3	1.69E+06	5.58E+06	8.83E+05	3.31	0.52	0.16
4	3.62E+06	3.97E+06	8.97E+05	1.1	0.25	0.23
5	1.28E+07	1.15E+07	6.96E+06	0.9	0.54	0.6
6	3.52E+06	3.38E+06	1.59E+06	0.96	0.45	0.47
7	3.86E+06	4.16E+06	5.51E+06	1.08	1.43	1.32
8	4.85E+06	3.34E+06	3.25E+06	0.69	0.67	0.97
9	2.84E+06	2.54E+06	1.25E+06	0.89	0.44	0.49
10	1.71E+07	3.97E+06	2.59E+06	0.23	0.15	0.65
11	1.81E+06	7.00E+05	9.69E+05	0.39	0.54	1.38
12	5.27E+06	3.86E+06	3.54E+06	0.73	0.67	0.92
13	4.42E+06	1.72E+06	1.42E+06	0.39	0.32	0.83
14	1.93E+06	1.59E+06	3.31E+06	0.83	1.72	2.08
15	3.31E+06	2.96E+06	1.25E+06	0.9	0.38	0.42
16	6.04E+06	3.63E+06	2.45E+06	0.6	0.4	0.67
17	4.11E+06	2.03E+06	1.74E+06	0.49	0.42	0.86
18	9.94E+06	1.17E+07	6.02E+06	1.18	0.61	0.52
19	4.72E+06	2.78E+06	1.95E+06	0.59	0.41	0.7
20	5.03E+06	4.03E+06	2.42E+06	0.8	0.48	0.6
21	1.15E+07	5.49E+06	1.29E+07	0.48	1.12	2.36
22	4.99E+06	9.10E+05	5.42E+06	0.18	1.08	5.95
23	3.60E+06	9.82E+05	2.14E+06	0.27	0.59	2.17
24	2.67E+06	1.69E+06	1.32E+06	0.63	0.49	0.78
25	9.80E+06	9.40E+06	2.45E+06	0.96	0.25	0.26
26	7.37E+06	5.32E+06	8.37E+06	0.72	1.14	1.57
27	8.19E+06	1.09E+07	1.45E+06	1.33	0.18	0.13
28	1.14E+07	3.41E+06	2.79E+06	0.3	0.24	0.82
29	5.78E+06	5.72E+05	6.89E+05	0.1	0.12	1.21
30	8.02E+06	3.14E+06	6.25E+05	0.39	0.08	0.2
			Mean	0.79	0.59	1.03

a) Quantity values are average values. The ratios are calculated between methods.

Table 11 . Quantity values for landmarked spots from sample D, IPG 6-11's^{a)}

Landmark	In-Tray Method			Active/Cup	Passive/Cup	Passive/Active
	Cup Quant.	Active Quant.	Passive Quant.			
1	2.98E+06	1.82E+06	3.48E+06	0.61	1.17	1.91
2	8.74E+06	7.53E+06	1.96E+07	0.86	2.24	2.6
3	3.43E+06	3.12E+06	4.42E+06	0.91	1.29	1.42
4	2.62E+06	5.28E+06	4.09E+06	2.02	1.56	0.78
5	3.44E+06	4.27E+06	5.05E+06	1.24	1.47	1.18
6	1.95E+06	4.95E+06	6.25E+06	2.54	3.21	1.26
7	2.31E+06	2.00E+06	6.07E+06	0.87	2.63	3.04
8	1.30E+06	1.73E+06	3.61E+06	1.33	2.78	2.09
9	5.03E+06	9.04E+06	1.86E+07	1.8	3.69	2.06
10	2.49E+06	2.71E+06	3.01E+06	1.09	1.21	1.11
11	1.71E+07	1.54E+06	1.76E+07	0.09	1.03	11.39
12	2.69E+06	1.15E+06	7.32E+06	0.43	2.72	6.39
13	8.03E+06	1.51E+06	3.28E+06	0.19	0.41	2.17
14	1.14E+07	3.56E+06	1.70E+07	0.31	1.5	4.78
15	8.81E+06	6.20E+06	9.60E+06	0.7	1.09	1.55
16	4.41E+06	1.08E+06	1.11E+06	0.25	0.25	1.03
17	1.58E+07	2.10E+07	2.71E+07	1.33	1.72	1.29
18	1.12E+07	2.78E+06	2.44E+07	0.25	2.17	8.78
19	5.46E+06	1.79E+06	7.46E+06	0.33	1.37	4.17
20	1.57E+06	1.30E+06	4.55E+06	0.83	2.91	3.5
21	2.57E+06	3.85E+05	6.08E+05	0.15	0.24	1.58
22	3.38E+06	1.22E+06	2.67E+06	0.36	0.79	2.19
23	1.05E+07	1.37E+07	6.16E+06	1.3	0.59	0.45
24	9.95E+06	7.61E+06	1.25E+07	0.76	1.25	1.64
25	2.90E+06	4.70E+06	4.29E+06	1.62	1.48	0.91
26	3.49E+06	3.75E+06	3.69E+06	1.08	1.06	0.98
27	5.06E+06	3.33E+06	3.24E+06	0.66	0.64	0.97
28	1.52E+07	6.09E+06	5.47E+06	0.4	0.36	0.9
29	5.14E+06	4.31E+06	3.64E+06	0.84	0.71	0.85
30	4.74E+06	2.85E+06	2.10E+06	0.6	0.44	0.74
31	4.00E+06	5.56E+05	2.67E+06	0.14	0.67	4.8
32	5.19E+06	7.57E+06	8.37E+06	1.46	1.61	1.1
33	1.02E+07	5.22E+06	3.23E+06	0.51	0.32	0.62
34	1.30E+07	1.25E+06	4.86E+05	0.1	0.04	0.39
35	8.81E+06	5.27E+06	4.20E+06	0.6	0.48	0.8
			Mean	0.82	1.34	2.33

a) Quantity values are average values. The ratios are calculated between methods.

CHAPTER 5

PROTEOMIC MAPPING OF THE EXTREMOPHILIC ARCHAEON, *SULFOLOBUS SOLFATARICUS* P2Chapter Summary

A proteomic map of *Sulfolobus solfataricus* P2, an extremophilic archaeon that grows optimally at 80°C and pH 3.2, was developed using high resolution two-dimensional gel electrophoresis (2-DE) and peptide mass fingerprinting. A differential solubilization technique was employed to subfractionate cellular proteins into aqueous tris-soluble and aqueous tris-insoluble fractions. Enhanced resolution was achieved by expanding crowded isoelectric focusing regions and by using a gradient gel for size separation. Up to 1300 spots were resolved on large format (20 X 20 cm) gels. A total of 867 protein spots (659 tris-soluble(aq) proteins and 208 tris-insoluble(aq) were mapped over IPG 3-10, 4-7, and 6-11, with second dimension gels made of 8-18% polyacrylamide. A total of 325 different gene products were represented by the 867 proteins, with 275 gene products being identified in the tris-soluble(aq) fractions and 100 gene products in the tris-insoluble(aq) portion. Fifty gene products were found on gels from both fractions. This protein mapping study was able to confirm the expression of many of the proteins involved in numerous metabolic pathways, transport, energy production and nucleic acid replication, translation, and transcription. Tris-soluble(aq) fractions contained cytosolic proteins which participate in energy metabolism and DNA replication, translation and transcription. Tris-insoluble(aq) fractions contained many

membrane associated proteins, including ABC transporters and ATP synthase. This is the first proteomic mapping of a member of the Crenarchaeota and provides a foundation to better understand the biochemical pathways and mechanisms employed by *Sulfolobus* to survive in its extreme environment.

Introduction

Discoveries of microorganisms capable of thriving in environmental extremes from boiling hot springs, deep-sea hydrothermal vents, acid mine drainages (as low as pH 0), to inside porous Antarctic rocks have created great interest in the scientific community. Environmental conditions which were once thought of as barriers to life are now seen as opportunities to better understand the strategies and mechanisms used to sustain life, and perhaps provide insight into the nature of early life on Earth. Organisms that survive in harsh conditions are known as extremophiles and different types of extremophiles are classified according to the physical or geochemical properties of their habitats (Hough and Danson, 1999). A physical property may include extremes in temperature, pressure, or radiation while geochemical properties include extremes in pH, salinity, dessication, or redox potential (Rothschild et al, 2001). 16S Ribosomal-RNA sequencing and other studies have established that many of these extremophiles represent a unique domain of life known as Archaea (formerly Archaeobacteria), which are distinct from Bacteria (eubacteria) and Eucarya (eucaryotes). The complete genomic sequences of a number of extremophiles have been completed or are nearing completion

(<http://www.tigr.org/tigr-scripts/CMR2/CMRHomePage.spl>), providing a wealth of genomic information. Using the open reading frames, proteomic studies can now complement genomic findings by illuminating the protein products expressed by the microbes to cope with the environmental variables they are exposed to, as well as provide information on the post-translational modifications that modulate protein activity and subcellular location of those proteins within the cell.

Phylogenetic analyses of the Archaeal domain of life divide it into two lineages: Euryarchaeota and Crenarchaeota (DeLong and Pace, 2001; Hershberger et al, 1996; Pace, 2001). A third lineage, Korarchaeota, has also been proposed, but currently consists of only a few sequences and additional data is needed (Barns et al, 1996; Reysenbach et al, 2000). Euryarchaeota are comprised of a heterogeneous group of organisms which include the extreme halophiles and methanogens, the sulfur-dependent hyperthermophile, *Thermococcales*, and the thermophilic/acidophilic Thermoplasmales (Granoff and Webster, 1999). Crenarchaeota are more phenotypically homogeneous with members being hyperthermophiles and thermophiles (Hershberger et al, 1996). Crenarchaeota include the anaerobic hyperthermophile, *Thermoproteales*, and the hyperthermophilic/acidophilic obligate aerobe, *Sulfolobus*. Currently, Sulfolobales represent archaea with the most eucaryotic-type features as suggested by the presence of introns, the structure of their 50S rRNA, mechanisms of their genetic machinery, and the presence of proteins belonging to the histone deacetylase family (Seeger et al, 1993; Sensen et al, 1998; She et al, 2001).

The genus *Sulfolobus* was first described by Brock in 1972 as a sulfur-oxidizing “bacteria” containing lobed cocci, capable of growing at high temperature and low pH (Figure 21) (Brock et al, 1972). It is now known that this archaeon grows much faster on organic substrates than chemolithotrophically and that a variety of species and strains exist (Brock et al, 1972; Schafer, 1996). The genus most notably includes *S. solfataricus*, *S. acidocaldarius*, *S. brierleyi*, *S. shibatae*, *S. metallicus*, and *S. tokodaii*. *Sulfolobus* strains have been isolated from Yellowstone National Park (USA), El Salvador, New Zealand, Italy, Iceland, the West Indies, and Japan (Schafer 1996). The natural habitats of *Sulfolobus* are usually geothermal areas where elemental sulfur is abundant, especially in solfataric hot-springs. These springs can range in pH from 0.5-6 and have temperatures from 60 °C to nearly 100 °C, with optimum *Sulfolobus* growth occurring at 80 °C and pH 3 (Brock et al, 1972, Zillig et al, 1980). Currently, *S. solfataricus*, *S. acidocaldarius*, and *S. tokodaii* have been the most studied, because of their ability to grow aerobically and heterotrophically to high cell densities under simulated native conditions. More over, the complete genome of *S. solfataricus* P2 (2.99 Mbp, AE006641) and *S. tokodaii* strain 7 (2.69 Mbp, AB022438) have been published (Kawarabayasi et al, 2001 and She et al, 2001, respectively).

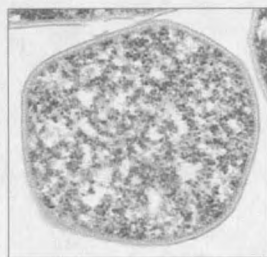


Figure 21. Electron microscope image of a cross-section of *S. solfataricus* P2.

The genome of *Sulfolobus solfataricus* P2 contains a total G+C content of 36% and encodes 2977 proteins, of which 33% have unknown function (hypothetical proteins). Based on amino acid sequence homology, 40% of the predicted proteins are archaeal specific and 12% are exclusive to bacterial protein matches, while only 2.3% are specific to eucaryotic protein matches. Twenty-three percent (700 ORFs) were homologous to both bacteria and eucarya (She et al, 2001). She, et. al., identified fifty-two different gene families from genomic data, with the largest members containing proteins involved in fatty-acid biosynthesis (acetyl-CoA synthetases), alcohol and other dehydrogenases, and ATP transporters (She et al, 2001). They also identified a total of 43 tRNA genes with 18 of them containing introns (She et al, 2001). Fourteen of the 18 introns were also found in other archaea (She et al, 2001), suggesting these introns are wide-spread in archaea.

To date, few investigations have used proteomics to identify proteins expressed by extremophiles. A global proteomic analysis was performed on *Methanococcus jannaschii* by Giometti et. al., in 2002, which identified 170 gene products on 2-D gels (Giometti et al, 2002). To our knowledge, this was previously the most extensive 2-DE mapping (which included spot identification) of an archaeon, until this study. Protein identifications in Giometti's publication were particularly sound since they used both peptide mass fingerprinting and peptide sequence tags (MS/MS) to identify the proteins. Giometti along with Adams and others, resolved approximately 500 proteins using 2-DE from the hyperthermophilic archaeon, *Pyrococcus furiosus*, but only identified 8 proteins out of the 500 (Giometti et al, 1995). Kawaguchi and Kuramitsu demonstrated the utility of classical proteomics on the bacterium, *Thermus thermophilus* HB8, by resolving 1200

proteins on 2-DE using a combination of narrow and broad range IPG gels (Kawaguchi, 1995), but only 22 spots were identified using N-terminal sequencing. Krufft and Wittmann-Liebold investigated the surfaces of intact ribosomal subunits from *Escherichia coli*, *Bacillus stearothermophilus*, and *Sulfolobus acidocaldarius* using 2-D gels (Krufft, 1991). Oberson, et. al., showed that the expression of heat shock proteins (HSP) were similar between the thermophilic fungus, *Chaetomium*, and the related mesophilic species *C. brasiliense* under stress-free conditions (Oberson et al, 1999); but that during stress, there was extensive variability in HSP expression, especially for the HSP60 family. Using 2-DE, Ding, et. al., uncovered the unique expression of multiple enzymes for *Methanosarcina thermophila* during methanogenesis when methanol was used in the culture medium as compared to acetate (Ding et al, 2002).

Other studies investigating the properties of proteins from extremophilic organisms have been primarily limited to proteins purified after cloning and expression in *Escherichia coli*. While conventional genetic and biochemical approaches are valuable ways to explore some of the properties of expressed proteins, especially from species that are hard to grow in the lab, proteomic methodologies are better able to address global questions about protein expression levels, variations in the mature form(s) of functional proteins, and as a way to identify those protein players most intimately associated with a particular stimuli.

The ability to cope with harsh environmental variables presumably requires the existence of unique cellular machinery. *Sulfolobus*, for example, maintains a cytosolic pH of 6.5 while growing in a pH 2-4 external environment and is able to take advantage

of the large pH gradient created across its plasma membrane (via ATP driven proton pumps) to generate energy for cellular processes (Schafer, 1996; Schafer, 1999; She et al, 2001). Application of proteomic techniques to studying extremophiles will certainly add to our understanding of the many diverse cellular mechanisms used by extremophiles to survive their habitats and may even yield some new discoveries. Proteomic studies have the potential to provide us with a better understanding of evolution, the existence of novel metabolic pathways, the requirements for protein stability and function, and may reveal proteins with unique or previously unknown properties.

As a foundation for understanding the strategies and mechanisms used by *Sulfolobus* to cope with environmental stress, a baseline 2-D proteomic map of *S. solfataricus* P2 was produced for cells grown under optimum laboratory conditions. A differential solubilization technique was used to subfractionate cellular proteins into tris-soluble(aq) and tris-insoluble(aq) fractions to reduce sample complexity and provide some evidence of subcellular location. Tris-soluble(aq) fractions presumable contain cytosolic proteins, while tris-insoluble(aq) fractions contain membrane associated proteins. Resolution of the 2-D gels was enhanced by using a combination of nonlinear broad-range and overlapping short-range IPG strips as well as by using 8-18% polyacrylamide gels for SDS-PAGE. After spot detection and imaging, identification of the resolved proteins was accomplished by peptide mass fingerprinting. We hope that this initial proteomic investigation of *S. solfataricus* P2 sets the stage for future expression profiling studies of this organism so that we may better understand this extremophilic microbe, especially in deciphering the biochemical pathways it employs.

Electrophoresis Equipment and Chemicals

All material and supplies were the same as described in Chapter 4, with the following additions. Methyl-modified trypsin was obtained from Promega (Madison, WI, USA) and α -Cyano-4-hydroxycinnamic acid, 99%, was purchased from Aldrich Chem. Co. (Milw., WI, USA).

Cell Culture and Harvesting

Sulfolobus solfataricus P2 (DSM 1617) was grown aerobically at 80°C and pH 3.2 in modified 9K medium (Schleper et al, 1994), without additional vitamins and containing 0.1% tryptone (wt/v). Cultures were grown in long-neck Erlenmeyer flasks in a shaking oil bath incubator for five to seven days to a late log-phase density of 0.6 (OD_{650}). Cells were isolated by centrifugation at 2000 X g at 25°C for 15 min and the cell pellets resuspended and washed twice in 50 mM Tris-HCl buffer, pH 7.4. The washed pellets were lyophilized overnight.

Differential Solubilization of Proteins and Preparation for IEF

Extraction of Tris-Soluble(aq) Proteins

Ten milligrams of dried cells were resuspended in 1 mL of 50 mM Tris-HCl, pH 7.4 containing 1 mM PMSF and sonicated for six cycles of 30 seconds duration with

intervening cooling on ice, using a Branson Sonifier 450 with titanium probe-tip, 50% duty cycle, and power output control at 2. After sonication, the lysed cells were incubated with 100 ng of DNAase and 100 ng RNAase for 1 h at room temperature. Next, the material was spun at 2000 X g for 5 min. to remove unbroken cells and the cloudy supernatant collected. The cloudy supernatant was then spun at 20,800 X g for 1 h, at 15 °C, using a benchtop Eppendorf Centrifuge (Model 5417) to separate the tris-soluble fraction from the tris-insoluble material. 500 µL of the tris-soluble supernatant was added to solid urea and CHAPS so that the final concentration was 8 M Urea, 2% CHAPS, and 50 mM Tris-HCl. Additionally, the solution was made to 4 mM tributylphosphine (TBP) and 0.2% carrier ampholytes 3/10 and allowed to stand for at least 2 h.

Extraction of Tris-Insoluble (aq) Proteins

The tris-insoluble(aq) pellet was washed three times in 50mM Tris-HCl buffer, pH 7.4, containing 1 mM PMSF. After each washing step the solution was spun at 14,000 X g for 10 min. using a benchtop Eppendorf centrifuge and the pellet collected for the next washing cycle. To the washed pellet, 400 µL of solution containing 5 M Urea, 1.5 M Thiourea, 3% CHAPS, 1% ASB-14, 50 mM Tris-HCl, 4 mM TBP, and 0.2% carrier ampholytes 3/10 was added and the mixture allowed to stand for at least 4 h. Finally, the sample was spun at 14,000 X g for 10 min. to remove any remaining insoluble material

and the supernatant used for 2-DE (defined as the detergent and chaotrope soluble extract, DCS).

Two-Dimensional Gel Electrophoresis

Both the tris-soluble(aq) and tris-insoluble(aq) fractions were resolved separately over three different immobilized pH gradients: broad-range NL 3-10, short-range 4-7, and IPG 6-11. IEF was performed using the Protean® IEF Cell and its 17 cm IEF tray. In order to fit the 18 cm Amersham Pharmacia strips into the IEF tray, the strips were cut to 17.5 cm, with the IPG 6-11's cut at the basic end, the IPG 4-7's cut at the acidic end, and the IPG NL 3-10's cut minimally at both ends. Tris-soluble(aq) protein concentrations (in TS solution) were based on Bradford assays, using bovine serum albumin as the standard, and 200 µg of protein was loaded onto each IPG strip. DCS fractions produced erroneous Bradford assay results and, therefore, 350 µL of DC solubilized solution obtained from the original 10 mg/mL solution of reconstituted cells was applied to an IPG strip to separate the tris-insoluble(aq) proteins.

Sample application by active rehydration was performed with IPG NL 3-10 and IPG 4-7 separations, while cup-loading was used to load protein onto IPG 6-11 strips, as previously described (Chapter 4, Barry et al, 2003a). All active rehydration applications were focused to 87,000 Vh using the following program: 50 V constant for 14 h, linear ramp to 300 V over 3.5 h, linear ramp to 1000 V over 2.5 h, linear ramp to 2500 V over 2.5 h, then 3500 V constant for 18 h, and 5000 V constant for 3.5 h. Cup-loading experiments were also focused to 87,000 Vh but used a slightly different prefocusing

phase: 100 V constant for 2.0 h, linear ramp to 300 V over 6.0 h, linear ramp to 1000 V over 2.5 h, linear ramp to 2500 V over 2.5 h, then 3500 V constant for 18 h, and 5000 V constant for 3.5 h.

After focusing, IPG strips were immediately equilibrated using 6 M urea, 50 mM Tris-HCl, pH 8.8, 30% glycerol, 4% SDS, and 2% DTT for 15 min followed by 6M urea, 50mM Tris-HCl, pH 8.8, 30% glycerol, 4% SDS, and 8% IAA for 15 min. The IPGs were then embedded in hot 0.5% agarose (70°C), containing bromophenol blue, on top of homemade 8-18% polyacrylamide gels (20 cm x 20 cm x 1.5 mm). SDS-PAGE was performed at 5 mA per gel for 2 h, followed by 20 mA per gel until the dye front reached the bottom of the gel.

Spot Detection and Image Analysis

Upon completion of SDS-PAGE, gels were fixed in 10% methanol and 7% acetic acid for 1 h with gentle agitation. The fixing solution was then removed and 300 mL of SYPRO Ruby added per gel and the gels stained overnight with agitation. Gels were then destained for three hours using 10% methanol and 7% acetic acid with one change of destaining solution. Two gels were fixed and stained per plastic container. Fluorescent scanning was performed using the Bio-Rad Molecular Imager™ FX at 200 µm pixel resolution, using a 532 nm laser, 1064 nm blocking excitation filter, and 555 nm long-pass emission filter. Image analysis and spot counting was performed using PDQuest 7.0 (Bio-Rad). To visualize the proteins for manual spot excision, the 2-D gels were then

stained with Coomassie G-250 for 24 h (Neuhoff et al, 1988; Neuhoff et al, 1990), followed by destaining with 1% acetic acid for 24 h to remove unbound dye, in order to visualize the proteins for manual spot excision.

In-Gel Proteolysis and Peptide Extraction

Protein spots were excised using a scalpel and a pair of tweezers and stored in 96-well microtiter plates. Coomassie stain was removed from the excised spots by three washings with 50% MeCN: 50% 50mM ammonium bicarbonate (pH 7.8) at 37 °C with gentle shaking. Gel pieces were then placed under vacuum for 20 min. using a Speed-vac. Typically 8 µL of 10-15ng/µL of methyl-modified trypsin in 25 mM ammonium bicarbonate, pH 7.8, was added to the dried gel piece and the solution allowed to absorb into the rehydrating gel. for 30 min., at room temperature. Any excess trypsin solution not absorbed was removed. Next, 10 µL of 25 mM ammonium bicarbonate buffer, pH 7.8, was added, an adhesive sealing strip applied to cover the samples, and the samples incubated at 37 °C for 16 h with gentle agitation. Peptide extraction was accomplished using a combination of sonication and passive diffusion. 8 µL of 50% MeCN containing 0.1% TFA was added to each gel piece and the pieces sonicated for 30 min. Following sonication, the gel pieces were placed on a benchtop for 2 hours before removing an aliquot for mass spectral analysis.

Peptide Mass Fingerprinting

Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI TOF MS) peptide mass fingerprinting (PMF) was accomplished using either the Micromass ToF-Spec IIE or Bruker Biflex III mass spectrometer. Aliquots of 0.5 μ L of proteolyzed protein sample were spotted onto a MALDI target-plate followed by 0.5 μ L of saturated matrix solution (α -Cyano-4-hydroxycinnamic acid in 50% MeCN, 0.1%TFA). Between 150-750 spectra were acquired and summed from each sample, depending on the intensity of the peptide signal and the automated spectra judging parameters used. Internal calibration of MS spectra was performed using three autodigestion peptides from the methyl-modified trypsin (842.5099, 1045.5642, and 2211.1046). Proteins were identified by using peptide digest mass mapping algorithms: either Mass Lynx 3.4 and ProteinProbe[®] (Micromass) or Biotools 2.0 and Mascot 1.8 (Bruker and Matrix Science, respectively). Experimental mass-to-charge (m/z) values for the peptides were compared to theoretically calculated m/z values of tryptic peptide digestions of *Sulfolobus solfataricus* P2 proteins (as determined from the sequenced genome). Database search results were inspected manually to assess proper protein identification. Identifications obtained using ProteinProbe are reported as the number of peptide matches and the percent of the protein identified, while identifications using Mascot are reported with a MOWSE score and the percent of the protein identified.

Results and Discussion

Tris-soluble(aq) proteins were resolved over IPG NL 3-10, 4-7 and 6-11 ranges and further separated on 8-18% SDS-PAGE gels. A description of the IEF buffers used in obtaining each gel is in this the legend of each gel figure. The urea/thiourea/CHAPS/ASB-14 buffer has been shown to maintain protein solubility better during IEF than the urea/CHAPS buffer, especially for membrane associated proteins (Rabilloud, 1999). In theory, tris-soluble(aq) proteins should not require strong solubilization buffers during IEF to obtain good gels, since cytosolic proteins are not as hydrophobic as other proteins. On the other hand, focused proteins have a net charge of zero and, hence, are at their minimal solubility and, thus, stronger buffers can help minimize protein precipitation.

Each IEF buffer was used to obtain separations of tris-soluble(aq) proteins on the IPG 3-10 strips as a way of comparing the quality of the gels produced, as can be seen in Figures 23 and 24. Both buffers performed equally well and the 2-D gel patterns were very similar. Each technique was able to produce essentially the same number of spots (1285 using the urea/CHAPS buffer, and 1310 using the urea/thiourea/CHAPS/ASB-14 buffer). Four hundred spots were plucked from the "urea/CHAPS" gel and three hundred plucked from the "urea/thiourea/CHAPS/ASB-14" gel, subjected to trypsin proteolysis, digested and analyzed with MALDI MS for subsequent protein identification. A total of 179 spots (45% of those plucked) were definitively identified in the gel using the urea/CHAPS buffer (Figure 23), and represented 113 distinct gene products. A total of 143 spots (48% of those plucked) were identified from the "urea/thiourea/CHAPS/ASB-

14" gel, representing 105 different gene products (Figure 24). In general, the percentage of proteins which could be definitively identified is a reflection of the post-gel analysis technique and should not be viewed as dependent on the buffers used. That being said, sixty of the 105 gene products identified from the "urea/thiourea/CHAPS/ASB-14" gel were not identified on the urea/CHAPS gel. This is most likely due to post-2-D gel analysis (i.e. trypsin proteolysis and MS analysis), although there is a possibility that a some proteins could have remained in solution better when using the urea/thiourea/CHAPS/ASB-14 as compared to urea/CHAPS during IEF. It should be noted that multiple spots were identified as containing the same protein. These proteins were isoforms and contained post-translational modifications, such as deamidation, phosphorylation, glycosylation, and acetylation. Additionally, some spots contained more than one gene product and was due to limits on the resolution that can be achieved with 2-D gels.

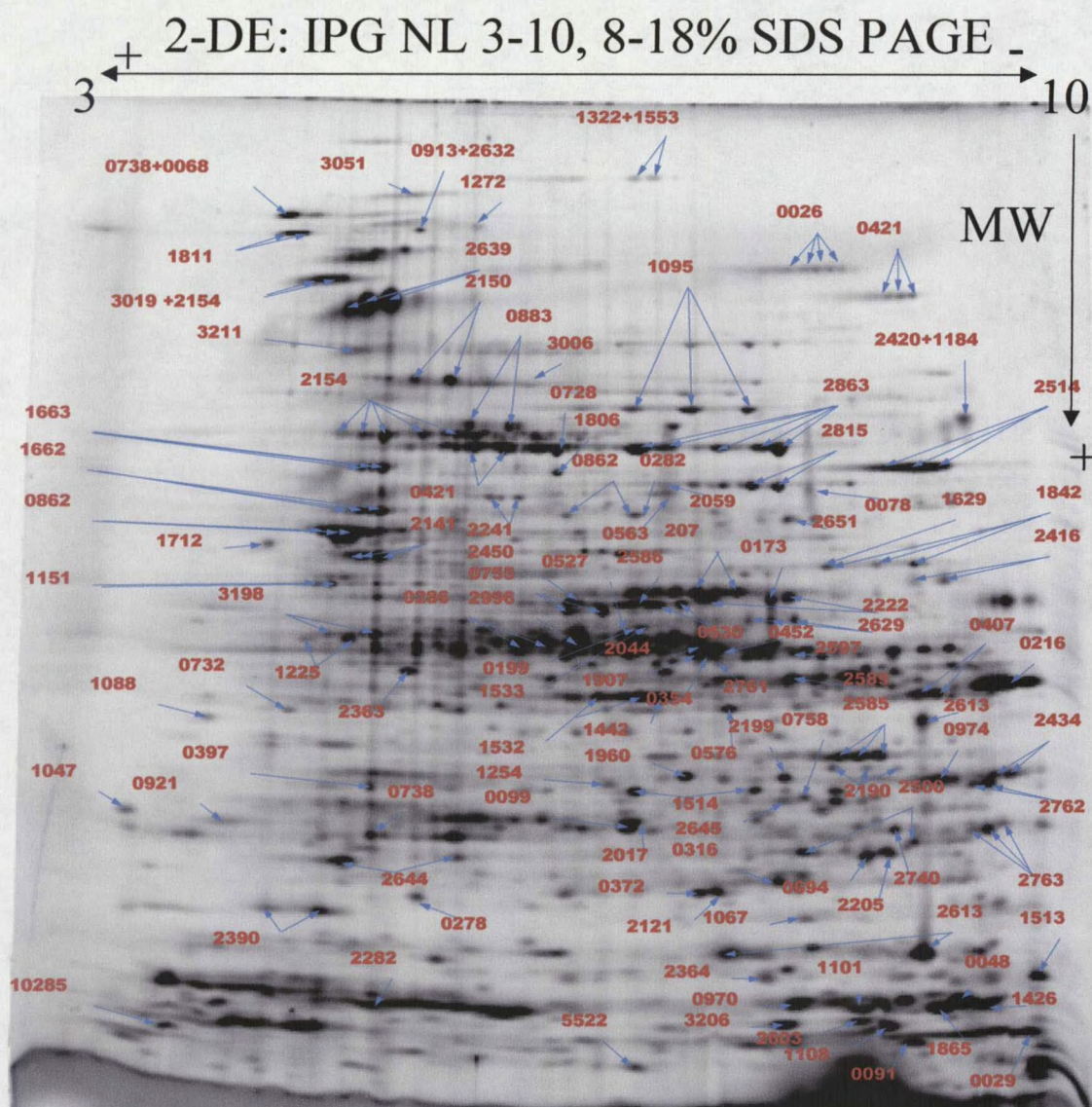


Figure 23. Two-dimensional gel electrophoresis (2-DE) of 200 μ g of tris-soluble(aq) proteins from *S. solfataricus* P2. Isoelectric focusing was performed using an IPG NL 3-10 strip. IEF buffer consisted of 8 M urea, 2% CHAPS, 2 mM TBP, in 50 mM Tris-HCl. SDS-PAGE was performed using an 8-18% polyacrylamide gel and Laemmli buffer. Gels were stained with SYPRO Ruby and imaged at 532 nm using the Molecular Imager FX (Bio-Rad). Proteins spots were excised and identified by PMF using ProteinProbe (Micromass). Spot labeling was performed using Delta2D software (Decodon). Spots are labeled according to their *S. solfataricus* P2 number SSOXXXX and identified in Table 12.

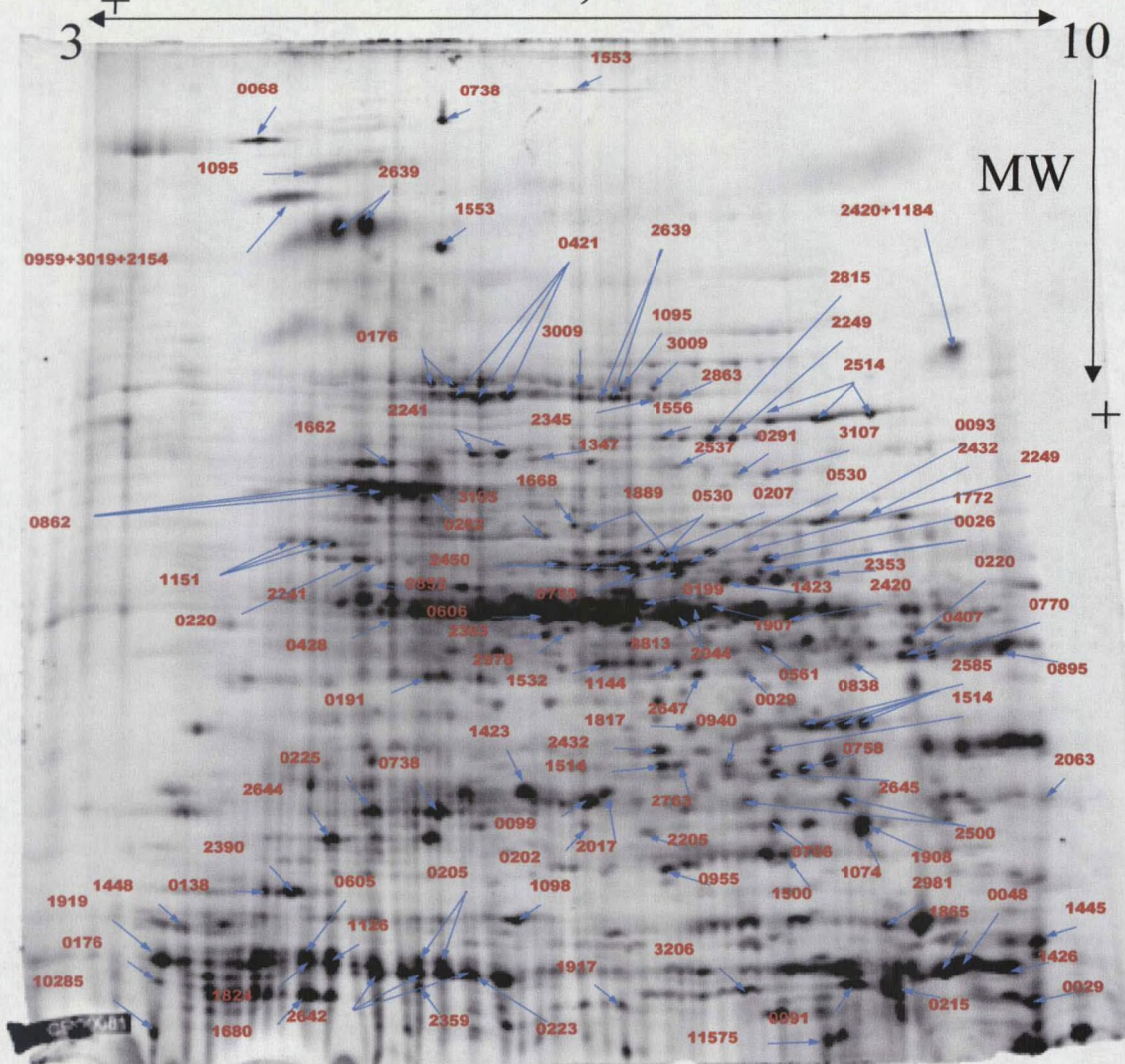


Figure 24. Two-dimensional gel electrophoresis (2-DE) of 200 µg of tris-soluble(aq) proteins from *Sulfolobus solfataricus* P2. Isoelectric focusing was performed using an IPG NL 3-10 strip. IEF buffer consisted of 5 M urea, 1.4 M thiourea, 3% CHAPS, 1% ASB-14, 2 mM TBP, in 50 mM Tris-HCl. SDS-PAGE was performed using an 8-18% polyacrylamide gel and Laemmli buffer. Gels were stained with SYPRO Ruby and imaged at 532 nm using the Molecular Imager FX (Bio-Rad). Proteins spots were excised and identified by PMF using Mowse scores from Mascot (Matrix Science). Spot labeling was performed using Delta2D software (Decodon). Spots are labeled according to their *S. solfataricus* P2 number SSOXXXXX and identified in Table 12.

The tris-soluble(aq) extracts were also separated using IPG 4-7 and 6-11 strips to enhance the resolution between the proteins clustered in those pH ranges and to provide additional evidence of the location and identification of the proteins. The enhanced resolution produces gel spots containing fewer proteins per spot which decreases the complexity of MS spectra and increases the likelihood of identifying proteins with database search algorithms. When using the narrower IPG ranges only the urea/CHAPs buffer was used during IEF of the tris-soluble(aq) fractions and 300 spots were plucked from the subsequent gels. When using the narrower IPG ranges to separate the tris-insoluble(aq) pellet only the urea/thiourea/CHAPS/ASB-14 buffer was used (since this fraction contains membrane associated proteins and this buffer was superior to extract membrane associated proteins from the tris-insoluble(aq) pellet).

Approximately 900 spots could be resolved for the tris-soluble(aq) proteins on gels obtained from the IPG 4-7 strips (Figure 25). A total of 187 spots were identified, representing 111 gene products. For the IPG 6-11 gels, approximately 600 spots were resolved and 150 proteins identified, representing 104 different gene products (Figure 26).

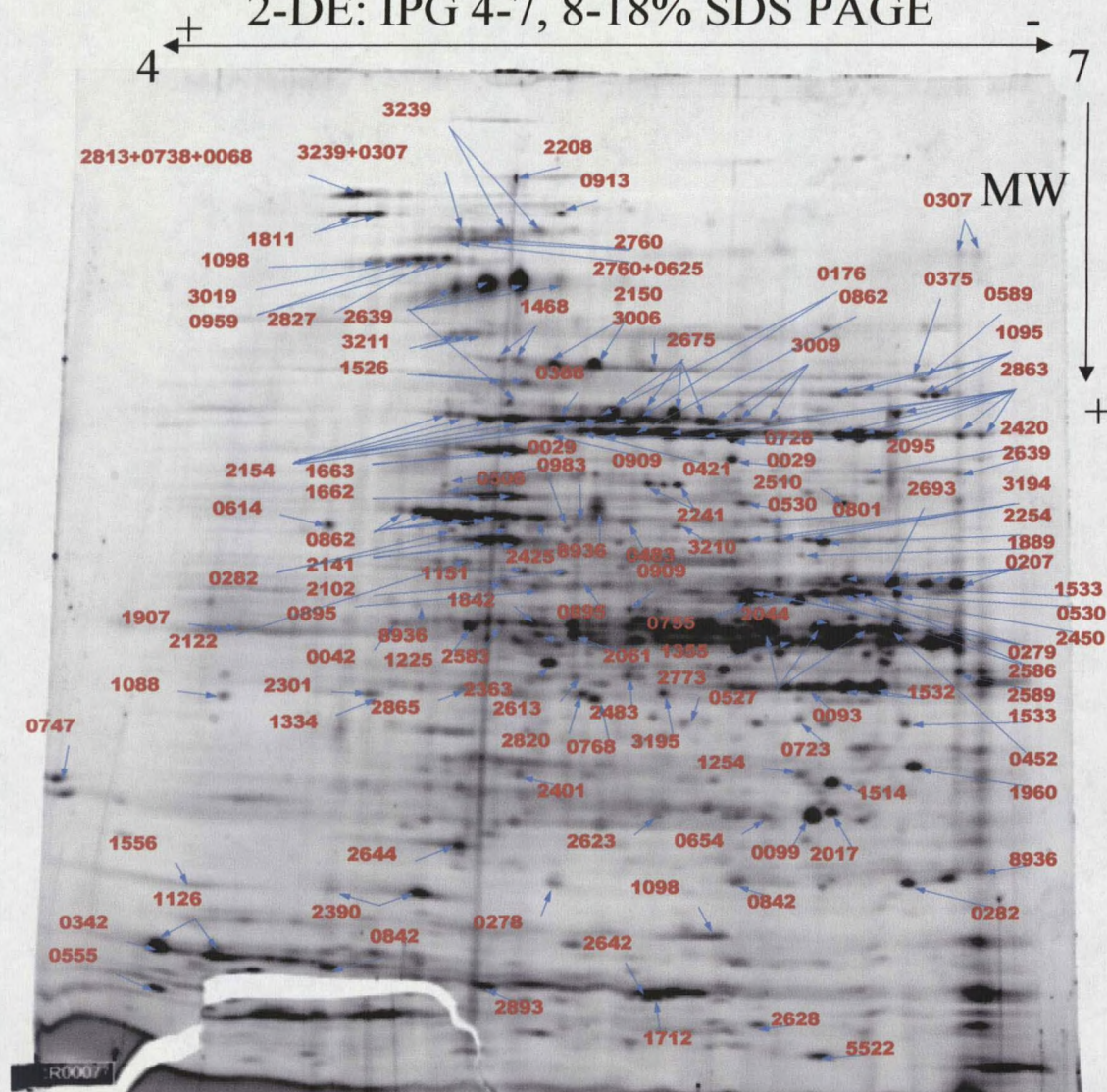


Figure 25. Two-dimensional gel electrophoresis (2-DE) of 200 µg of tris-soluble(aq) proteins from *Sulfolobus solfataricus* P2. Isoelectric focusing was performed using an IPG 4-7 strip. IEF buffer consisted of 8 M urea, 2% CHAPS, 2 mM TBP, in 50 mM Tris-HCl. SDS-PAGE was performed using an 8-18% polyacrylamide gel and Laemmli buffer. Gels were stained with SYPRO Ruby and imaged at 532 nm using the Molecular Imager FX (Bio-Rad). Proteins spots were identified by PMF using Mowse scores from Mascot (Matrix Science). Spot labeling was performed using Delta2D software (Decodon). Spots are labeled according to their *S. solfataricus* P2 number SSOXXXX and are listed in Table 12.

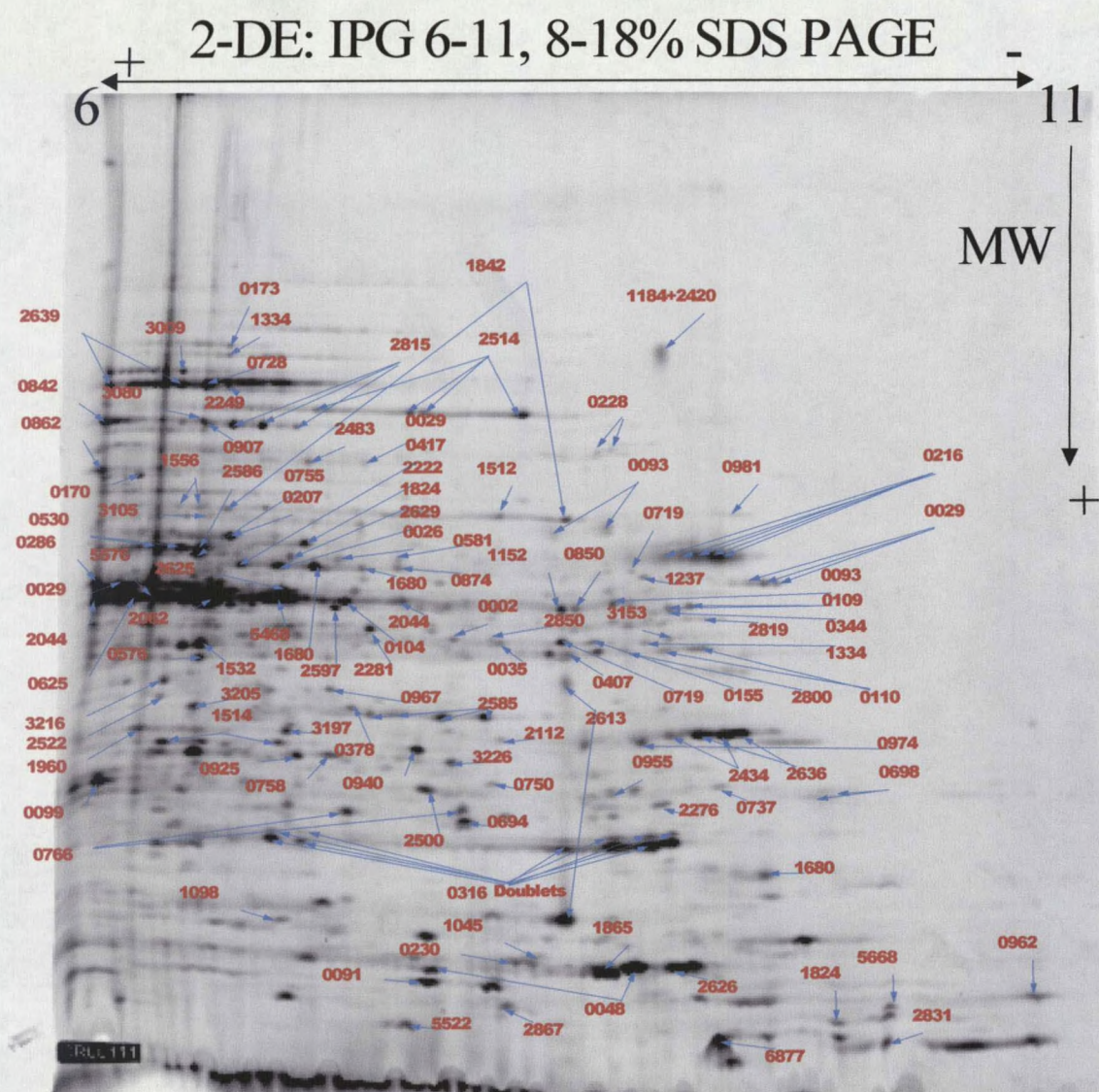


Figure 26. Two-dimensional gel electrophoresis (2-DE) of 200 μ g of tris-soluble(aq) proteins from *Sulfolobus solfataricus* P2. Isoelectric focusing was performed using an IPG 6-11 strip. IEF buffer consisted of 8 M urea, 2% CHAPS, 2 mM TBP, in 50 mM Tris-HCl.. SDS-PAGE was performed using an 8-18% polyacrylamide gel and Laemmli buffer. Gels were stained with SYPRO Ruby and imaged at 532 nm using the Molecular Imager FX (Bio-Rad). Proteins spots were identified by PMF using Mowse scores from Mascot (Matrix Science). Spot labeling was performed using Delta2D software (Decodon). Spots are labeled according to their *S. solfataricus* P2 number SSOXXXX and are listed in Table 12.

The same approach was used in analyzing the tris-insoluble(aq) fraction. The total number of proteins detected on all of the gels obtained from the aqueous tris-insoluble (TI) fractions was substantially lower than the numbers obtained from the aqueous tris-soluble fractions (TS). This was due to lower protein loading on the TI gels as compared to the TS gels and because the total number of membrane associated proteins is typically less than the number of cytosolic proteins. For comparison, approximately 250 spots were detected on the TI NL 3-10 gels versus 1285 detected on the TS NL 3-10 gel. Gel images obtained from the TI fraction are shown in Figures 27-29, with the identified spots labeled. A total of 208 spots could be identified, representing 100 different gene products.

Figure 27. Two-dimensional gel electrophoresis (2-DE) of tris-insoluble(aq) proteins from *Sulfolobus solfataricus* P2. Isoelectric focusing was performed using an IPG NL 3-10 strip. IEF buffer consisted of 5 M urea, 1.4 M thiourea, 3% CHAPS, 1% ASB-14, 2 mM TBP, in 50 mM Tris-HCl. SDS-PAGE was performed using an 8-18% polyacrylamide gel and Laemmli buffer. Gels were stained with SYPRO Ruby and imaged at 532 nm using the Molecular Imager FX (Bio-Rad). Proteins spots were identified by PMF using Mascot (Matrix Science). Spot labeling was performed using Delta2D software (Decodon). Spots are labeled according to their *S. solfataricus* P2 number SSOXXXX and are listed in Table 12.

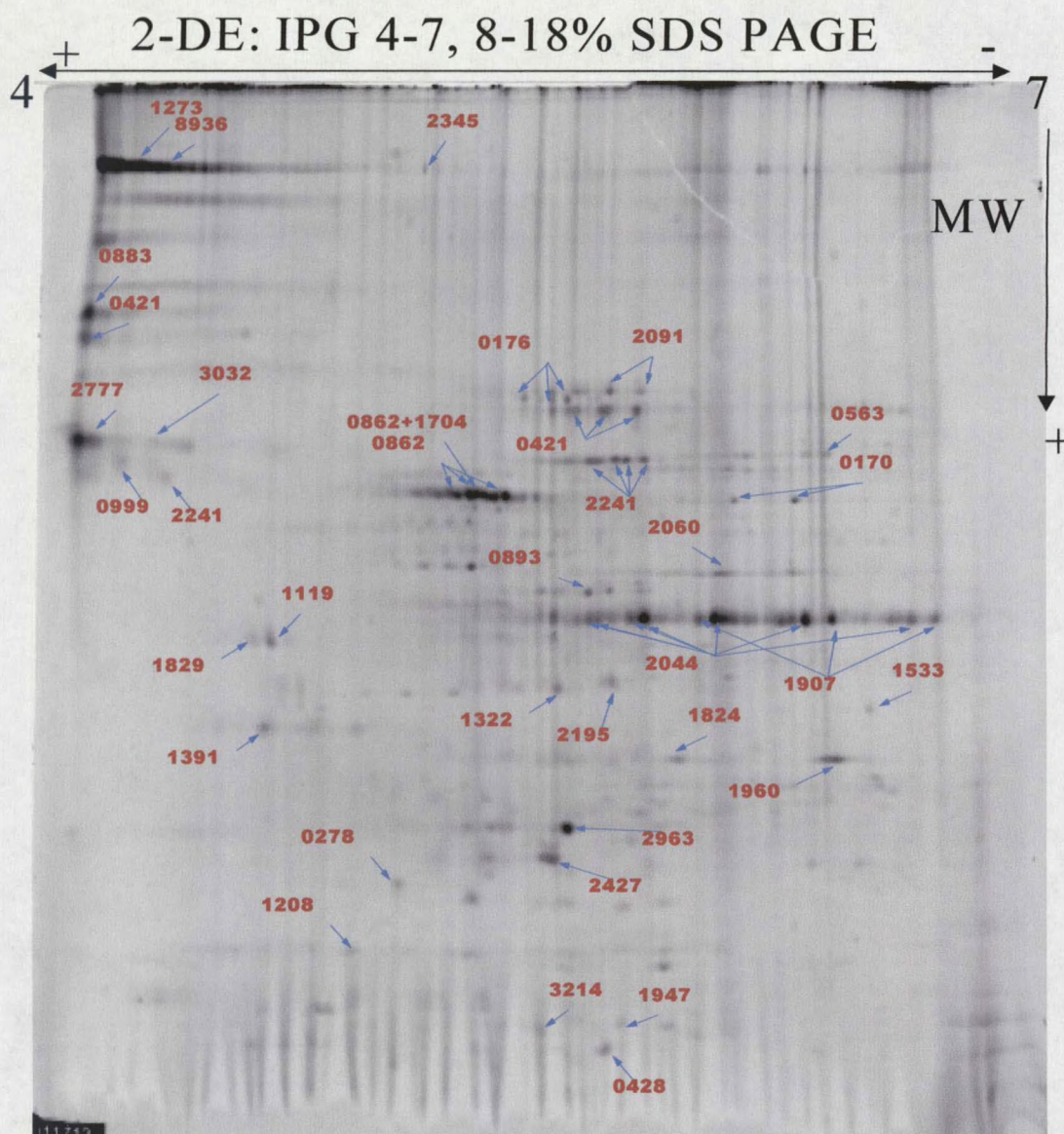


Figure 28. Two-dimensional gel electrophoresis (2-DE) of tris-insoluble(aq) proteins from *Sulfolobus solfataricus* P2. Isoelectric focusing was performed using an IPG 4-7 strip. IEF buffer consisted of 5 M urea, 1.4 M thiourea, 3% CHAPS, 1% ASB-14, 2 mM TBP, in 50 mM Tris-HCl. SDS-PAGE was performed using an 8-18% polyacrylamide gel and Laemmli buffer. Gels were stained with SYPRO Ruby and imaged at 532 nm using the Molecular Imager FX (Bio-Rad). Proteins spots were identified by PMF using Mascot (Matrix Science). Spot labeling was performed using Delta2D software (Decodon). Spots are labeled according to their *S. solfataricus* P2 number SSOXXXX and are listed in Table 12.

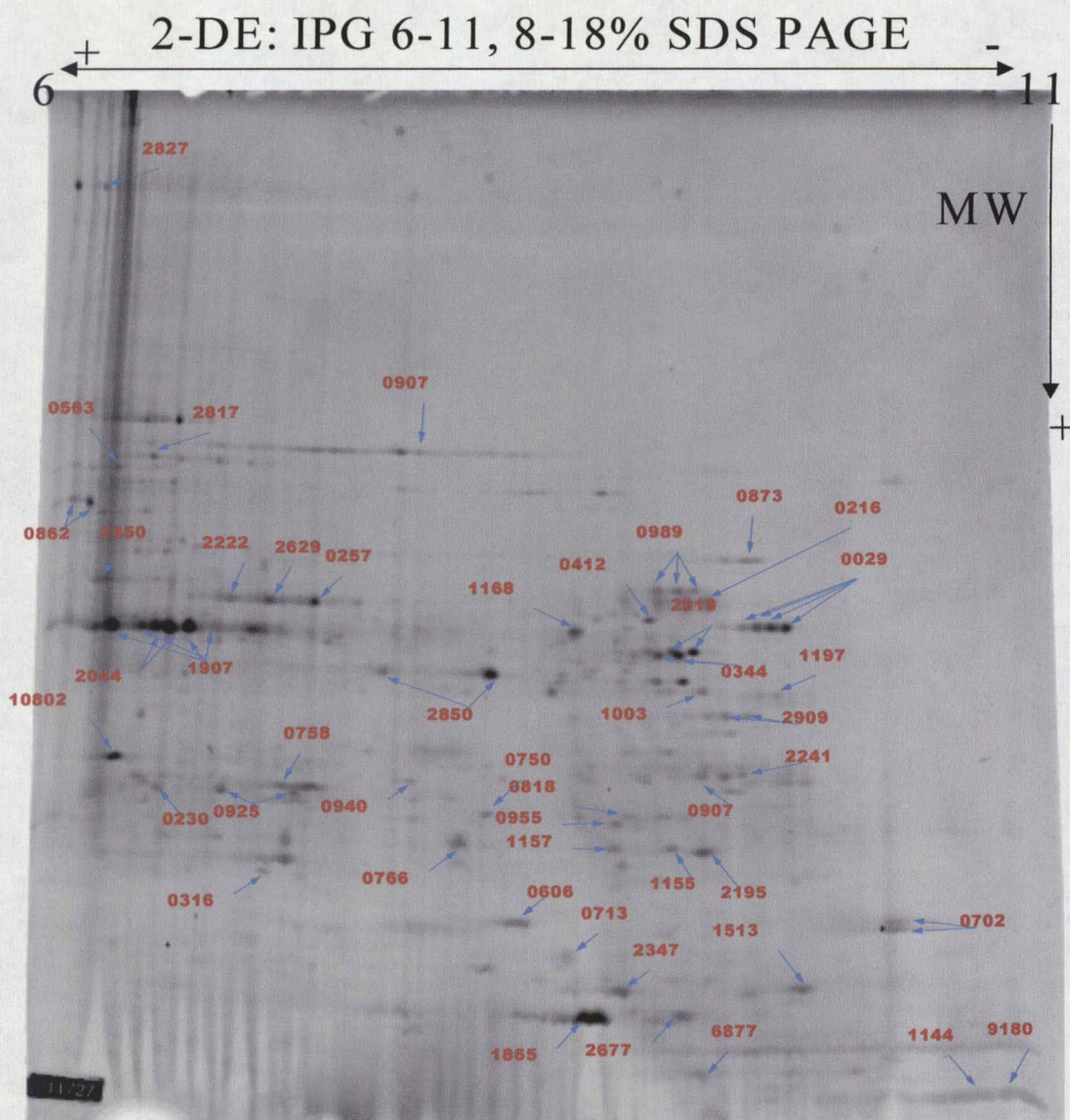


Figure 29. Two-dimensional gel electrophoresis (2-DE) of tris-insoluble(aq) proteins from *Sulfolobus solfataricus* P2. Isoelectric focusing was performed using an IPG 6-11 strip. IEF buffer consisted of 5 M urea, 1.4 M thiourea, 3% CHAPS, 1% ASB-14, 2 mM TBP, in 50 mM Tris-HCl. SDS-PAGE was performed using an 8-18% polyacrylamide gel and Laemmli buffer. Gels were stained with SYPRO Ruby and imaged at 532 nm using the Molecular Imager FX (Bio-Rad). Proteins spots were identified by PMF using Mascot (Matrix Science). Spot labeling was performed using Delta2D software (Decodon). Spots are labeled according to their *S. solfataricus* P2 number SSOXXXX and are listed in Table 12.

A listing of the 325 gene products identified from the various 2-D gels is shown in Table 12. The proteins are listed according to their *S. solfataricus* P2 ORF number (SSO #) along with the gels in which they were identified. If the protein was identified on a particular gel, the number of peptides matched and/or the MOWSE score of the match is listed along with the percentage of the protein identified from that match. Spots which were determined to be isoforms of the same protein are only listed once, with the highest matching score found among those isoforms. SSO# numbers which appear twice indicate that it was found in one or both of the fractions, as color coded. Proteins which were identified from the tris-soluble(aq) fractions only have their SSO # colored black. Those gene products found in the tris-insoluble(aq) fraction only are shown with SSO # in red font. Fifty gene products were found on gels from both fractions and their SSO #'s are colored blue. The extent of overlapping was minimized by washing the tris-insoluble(aq) fraction three times with tris-HCl buffer (see methods section). The efficiency of the washing is evidenced by noticing that proteins which were detected with the highest abundance on the TS gels (e.g. SSO2639) were not detected on gels from the TI fractions. This suggests that many of the proteins which were found in both fractions are proteins which interact with the plasma membrane. More information about the proteins listed in Table 12 can be obtained at:

<http://www-archbac.u-psud.fr/projects/sulfolobus/>. A virtual 2-D gel consisting of all 2977 ORFs can also be found on the above Sulfolobus website (under the pI and MW link).

Overall, a total of 867 P2 spots were identified across all of the gels, with 659 tris-soluble(aq) proteins being mapped and 208 tris-insoluble(aq) proteins mapped. Of the 867, 325 different gene products are represented. 50 gene products were unique to the tris-insoluble(aq) fraction and 50 were found in both the tris-soluble(aq) and tris-insoluble(aq) fractions. 275 proteins were found only in the tris-soluble(aq) fraction.

Typically one expects tris-soluble(aq) proteins to primarily consist of cytosolic proteins and tris-insoluble(aq) fractions to contain membrane associated proteins. This was generally confirmed by our analyses. Table 12 shows that cytosolic proteins, such as those enzymes involved in different metabolic pathways and those that function in DNA replication, translation and transcription, were found in the tris-soluble(aq) fractions. Gels obtained from tris-insoluble(aq) fractions contained many proteins known to be membrane associated, including: ATPases, ATP synthetases, ABC transporters, and an Erythrocyte 7 TM homologous integral membrane protein (SSO2195).

A closer inspection the types of proteins identified shows that each of the major cellular processes were represented as expected. A short list of these proteins is presented in the next few paragraphs. Two different A-type (A_0A_1) ATPases, which are common to all three domains of life were identified (SSO561 and SSO563). Protein turnover is represented by two intercellular proteasome subunit proteins (SSO0738 and SSO0766) and two tricorn protease cofactors (SSO2154 and SSO2675). The proteins responsible for intracellular synthesis of trehalose (SSO2095), glycogen degradation (SSO0988), glycogen synthesis (SSO0989) and a glycosidase used in cytoplasmic conversion of polysaccharides to monosaccharides (SSO3019) were also identified.

The central metabolic pathways of glycolysis and the citric acid cycle were well represented. Enzymes representing the non-phosphorylating Entner-Doudoroff pathway included enolase (2-phosphoglycerate dehydratase, SSO0913), pyruvate kinase (SSO0981), and 2-keto-3-deoxy gluconate aldolase (SSO3197). The citric acid cycle associated proteins included aconitate hydratase (SSO1095), the classical succinate dehydrogenase (SSO2359) (She et al, 2001), succinyl-CoA synthetase (SSO2483), citrate synthase (SSO2589), 2-oxoacid-ferredoxin oxidoreductase (SSO2815), and acetyl-CoA synthetase (SSO2863).

A protein homologous to sulfite reductase, that catalyzes the oxidation of sulfur to sulfate, was present (SSO2909). Proteins which are intimately involved in replication, transcription, and translation were also identified: a bacterial-like DNA primase (SSO0079), a fibrillarin-like pre-rRNA processing protein (SSO0940), a nucleoside diphosphate kinase (SSO0230), DNA topoisomerase I (SSO0907), DNA-directed RNA polymerases (SSO0225 and 0223), a protein used for DNA structural maintenance (SSO2249), a subunit for Glu-tRNA amidotransferase (SSO2122), three DNA binding proteins (SSO6877, SSO9180, and SSO2364), and an ATP-dependent RNA helicase was identified (SSO1889).

Many stress-responsive proteins were also identified. These include: the heat shock proteins (chaperones), HSP 55 (SSO0862) and HSP 20 (SSO2427); oxidative stress related proteins, superoxide dismutase (SSO0316), thioredoxin reductase (SSO2416), oxidoreductase (SSO2629), peroxiredoxin (a bacterioferritin, SSO2613), and rubrerythrin (SSO2642).

Some surprising proteins identified which may be of further interest include: glyoxylate oxidase (SSO3161), eight of the thirteen carbon monoxide dehydrogenases encoded by the P2 genome (cutA, B and C variety, EC 1.2.99.2, homologous with 4-hydroxybenzoyl-CoA reductase, xanthine dehydrogenase, and aldehyde dehydrogenase, respectively), a thiosulfate transferase (SSO1817), a sulfolipid biosynthesis protein (SSO2583), N-methylhydantoinase A and B (which catalyze the formation of sarcosine, SSO1662 and SSO1663), a NAD specific glutamate dehydrogenase (non-energy metabolism related, SSO1907 and SSO2044), a FKBP-type rotamase (a chaperone that catalyzes the conversion from cis-to-trans protein folding, SSO0758), a geranylgeranyl pyrophosphate synthetase and a geranylgeranyl hydrogenase protein (SSO2345 and 2353, involved in prenylation of proteins), and two carboxypeptidases (one hypothetical, SSO3105, and one previously characterized to be stable to 85 °C, SSO1355 (Colombo et al, 1992)).

An example of the type of MALDI MS data used to provide a peptide peak list and the corresponding MOWSE search results is shown in Figures 30 and 31 for phosphoenolpyruvate carboxykinase (PEPCK, SSO2537, EC 4.1.1.32). PEPCK is a key enzyme in gluconeogenesis, and catalyzes the conversion of oxaloacetate to phosphoenolpyruvate. This enzyme shares very high amino acid sequence homology to PEPCK proteins from many genera including, *Pyrococcus furiosus* (an E-value of e^{-166} , identities of 49% with 310/624, positives of 66% from 418/624 matches, with gaps of 3%, 23/624), *Giardia lamblia* (E-value of e^{-162}), and *Chlamydia trachomatis* (with an E-value of $4e^{-80}$), to name just three.

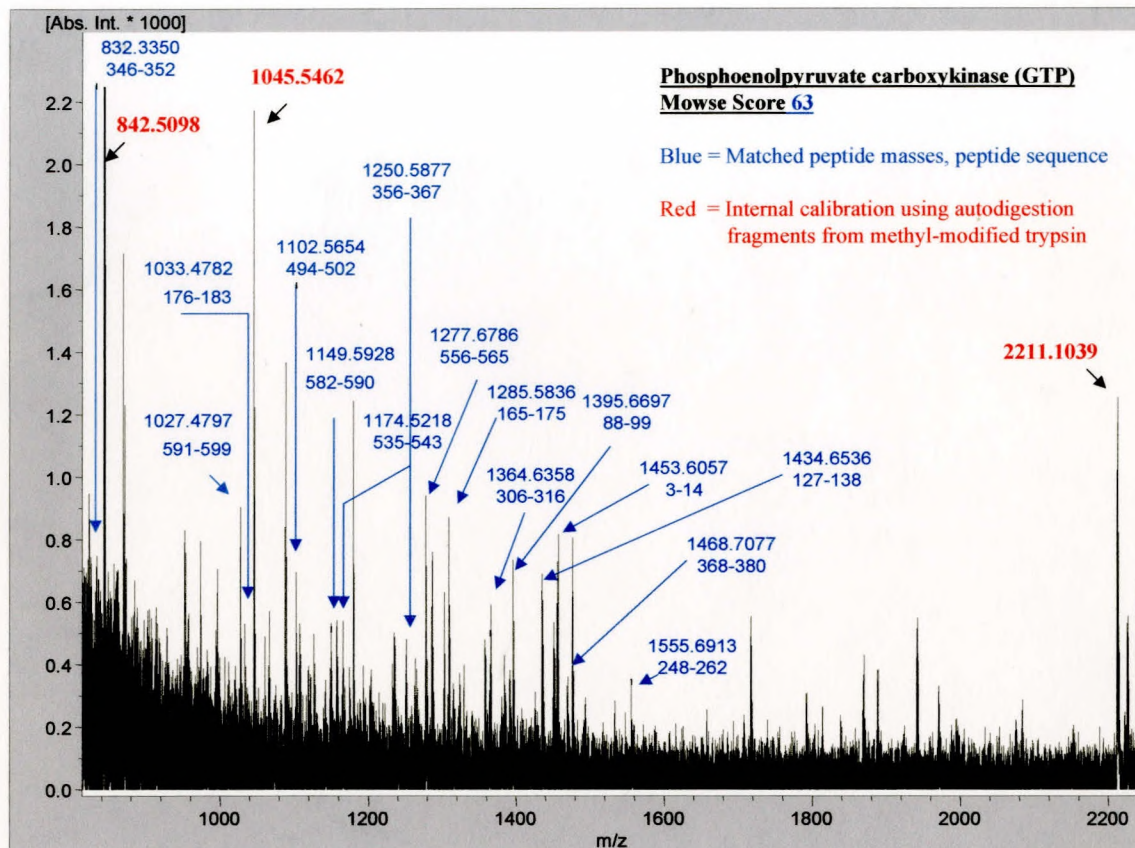


Figure 30. MALDI MS spectrum (Bruker Biflex III) that identified phosphoenolpyruvate carboxykinase (GTP), SSO2537. For high mass accuracy, internal calibration was performed using autodigestion fragments from methyl-modified trypsin. A peptide peaklist was generated and used for database searching (BioTools and Mascot). A MOWSE score of 63 was obtained providing > 95% confidence in the match, see Figure 31.

Phosphoenolpyruvate carboxykinase (GTP), SSO2537

10	20	30	40	50	60	70	80
MKSSLDLFLNN	FIQRDAVKKL	ESINNYPLIE	FLNSVVKLCE	PDSVYLITGS	DEEKEYIRKK	ALESKEEIRL	QTSGHTIHFD
90	100	110	120	130	140	150	160
HPLDQARARE	DTFILSDTKI	FYVNTKPRNE	GLSEMLTLK	GSMRGREMYV	GFYSLGPRNS	PFQILAVQVT	DSPYVIHSEN
170	180	190	200	210	220	230	240
ILYRIAFEDF	SNNTKFLRFV	HSKGEPDIKK	RRIMIDLANN	TVYSVNTTYA	GNSVGLKKLA	LRLTIMKAVE	EGWLSEHMAI
250	260	270	280	290	300	310	320
VGFNQDKGIH	YFTASFPSGS	GKTSTSMIGN	LISDDLAFIR	EFEGLPKAVN	PEIGVFGIIQ	GINARDDPII	WEVLHKPGEV
330	340	350	360	370	380	390	400
IFSNVLMTED	GDVYWEGSEL	PKPERGYNHE	GRWDRESGKP	ASHPNARFTV	PLTSFSNLDK	NWDNPNGVII	DGIIFGVRDY
410	420	430	440	450	460	470	480
STLVFVVEAF	SWSHGVITIG	ASMESARTSA	VIGKSDELEF	NPMAILDPMP	ISLSRYLRNY	LNFGKRLRKS	PKIFGFNYFL
490	500	510	520	530	540	550	560
KDDNKFLNSK	EDKKVWVSWA	VKRVEETANA	IYTPIGLIPF	YEDLKALPKR	VLGKEYGKEE	YEKQFTIKLR	KYLEKTERII
570	580	590	600	610			
EIYLKFEDIP	SEVINELKMQ	KERIVDYINK	YGDSVSPFRL	EKD			

Mowse Score 63
RMS error 52 ppm

15 peptide matches,
26% Coverage
pI = 6.77,
MW = 69,056 Da

Mass Mr	Dev.	Range	P	Sequence
831.3272	-0.0351	346 - 352	0	GYNHEGR
1026.4719	-0.0051	591 - 599	0	YGDSVSPFR
1032.4704	-0.1165	176 - 183	1	FLRFVHSK
1101.5576	-0.0759	494 - 502	1	KVWVSWAVK
1148.5850	-0.0339	582 - 590	1	ERIVDYINK
1173.5139	-0.0050	535 - 543	1	EYGKEEYK
1249.5799	-0.0364	356 - 367	1	ESGKPASHPNAR
1276.6708	-0.0683	556 - 565	1	TERIIIEIYLK
1284.5757	-0.0228	165 - 175	0	IAFEDFSNNTK
1363.6280	-0.0856	306 - 316	0	DDPIIWEVLHK
1394.6619	-0.0422	88 - 99	1	AREDTFILSDTK
1433.6458	-0.0191	127 - 138	0	EMYVGFYSLGPR
1452.5979	-0.1382	3 - 14	0	SSLDLNNFIQR
1467.6999	-0.0610	368 - 380	0	FTVPLTSFSNLDK
1554.6834	-0.0632	248 - 262	0	GIHYFTASFPSGSGK

Figure 31. P2 database search results for the spectrum in Figure 30. The amino acid sequence for PEPCK and the peptides matched (top). A MOWSE score of 63 was obtained and represented 15 experimental peptide matches (bottom) to an *in silico* tryptic digest of SSO2537. The average root-mean-square (RMS) of the matches was 52 ppm. The 15 peptides matched represent 26% of the amino acid sequence of the protein identified.

Concluding Remarks

Sulfobales are ubiquitous microorganisms that are found in extremely thermophilic (70-90°C) and acidophilic (pH 0.6-6) environments. Until recently, the idea that life could exist in such inhospitable places was unimaginable, but somehow this organism has developed a repertoire of cellular machinery that enhances its survival. At first glance, hyperthermophilic archaeons seem to have many of the traits one would have expected of ancient microorganisms when they inhabited primeval Earth. Surprisingly, genomic studies have revealed that archaeons, like *Sulfolobus*, use eucaryotic-type mechanisms for DNA replication, translation, and transcription, but bacterial-type biochemical pathways for metabolism (Cohen et al, 2003; Dionne et al, 2003; Schafer, 1996; White, 2000). Unfortunately, our fundamental understanding of hyperthermophilic archaea has been limited because of the lack of genetic tools to study gene expression and proteomic methods to identify gene products, map metabolic pathways, and track protein expression.

This is the first proteomic mapping published for any *Crenarchaeota* and establishes a technique that can be used to track and identify the mature protein forms used by *Sulfolobus* to cope with a wide range of environmental variables. Our technique resolved ~1300 spots on 2-D gels and we identified 867 spots, representing 325 different gene products (or 11% of the ORFs encoded by the P2 genome). Most of our current understanding of the biochemical pathways employed by *Sulfolobus* have come from the collective studies involving many different species of *Sulfolobus* and there have been

many confusing results. The techniques developed here will allow us to study the proteome from one species of *Sulfolobus*, thereby obviating any conflicting results derived from multiple species. We now have a method, for example, to better investigate the gluconeogenic and pentose phosphate pathways used by *Sulfolobus*, which is currently unknown. We have identified interesting proteins which reflect as yet unknown properties or pathways which invites further investigation, including: carbon monoxide dehydrogenase (CODH), geranylgeranyl synthetase, and phosphoenolpyruvate carboxykinase.

It is hoped that the methods and results published from this study will generate additional research to further explain the function(s) of some of the proteins identified and better illuminate the metabolic pathways used by *Sulfolobus solfataricus*.

Additionally, the methods developed here can be used to compliment genomic studies which are currently underway to develop gene transfer systems to study the gene expression systems for *Sulfolobus* (Aagaard et al, 1996; Aravalli and Garrett, 1997; Contursi et al, 2003; Jonscheit et al, 2003).

Table 12.

P2 Tris-Soluble and Tris-Insoluble Proteins Identified by *S. solfataricus* ORF Number, SSO

SSO# Column				IPG Column							
Black SSO# found only in the tris-soluble fraction				Black = ProteinProbe number of peptides found matching							
Red SSO # found only in tris-insoluble fraction				Red = Mascot Mowse matching score, 90% confidence or greater							
Blue SSO# found in both tris-soluble and tris-insoluble fractions				Blue = Mascot Mowse matching score, 95% confidence or greater							
Pct. Column = percentage of the protein identified with MS											
Type = protein functional category											
AAB: Amino Acid Biosynthesis						IS: Insertion Element					
CEM: Cell Envelope Membrane						Nuc: Nucleotide Synthesis/Metabolism					
CIM: Central Intermediary Metabolism						PPM: Proteases and Protein Modifications					
CF: Cofactor Biosynthesis						TrL: Translation					
CP: Cellular Processes						TrR: Transcription and Regulation					
EM: Energy Metabolism						TrSP: Transport					
LM: Lipid Metabolism						RR: Replication and Repair					
HP Hypothetical Protein						Unc: Uncategorized					
S.solf#	IPG 3-10	Pct	IPG 4-7	Pct	IPG 6-11	Pct	MW	pI	Type	Function	
SSO0001	44	26	—	—	—	—	23239	6.67	HP	Hypothetical protein	
SSO0002	—	—	—	—	54	21	43745	8.33	CF	(thiD-1) Phosphomethylpyrimidine kinase	
SSO0026	8/77	33/40	—	—	85	32	45538	7.63	Nuc	(purT) Phosphoribosylglycinamide formyltransferase 2	
SSO0029	15/54	30/26	86	45	150	38	43433	10.21	HP	Conserved hypothetical protein	
SSO0029	77	18	—	—	182	54	43433	10.21	HP	Conserved hypothetical protein	
SSO0035	—	—	—	—	58	41	12956	8.93	HP	Conserved hypothetical protein	
SSO0042	—	—	50	17	—	—	25071	11.36	ISE	Second ORF in transposon ISC1234	
SSO0048	10/72	32/46	—	—	66	48	14329	8.62	TrR	Transcriptional regulator, putative	
SSO0068	5/40	29/49	27	33	—	—	16152	10.91	TRL	(rps9AB) SSU ribosomal protein S9AB	
SSO0068	33	25	—	—	—	—	16152	10.91	TRL	(rps9AB) SSU ribosomal protein S9AB	
SSO0078	17	45	—	—	—	—	41819	8.34	TRL	(tyrS) Tyrosyl-tRNA synthetase	
SSO0079	93	40	—	—	—	—	45409	5.00	RR	(dnaG) Bacterial-like DNA primase	
SSO0091	8/70	17/30	—	—	47	24	14052	8.07	TRL	(rpl7AE) LSU ribosomal protein L7AE	
SSO0093	59	20	47	18	61	20	66805	8.91	TRL	(gltX) Glutamyl-tRNA synthetase	
SSO0093	64	20	—	—	—	—	66805	8.91	TRL	(gltX) Glutamyl-tRNA synthetase	
SSO0099	9/62	53/42	78	62	99	72	25108	6.24	HP	Conserved hypothetical protein	
SSO0104	—	—	—	—	96	28	45571	7.82	AAB	(aspB-1) Aspartate aminotransferase	
SSO0109	—	—	—	—	67	36	26580	9.58	HP	Conserved hypothetical protein	
SSO0109	34	16	—	—	—	—	26580	9.58	HP	Conserved hypothetical protein	
SSO0110	—	—	—	—	106	44	33521	9.91	HP	Conserved hypothetical protein	
SSO0138	33	24	—	—	—	—	14002	9.26	HP	Hypothetical protein	
SSO0155	—	—	—	—	91	49	38685	8.97	AAB	(argC) N-acetyl-gamma-glutamyl-phosphate reductase	
SSO0167	47	12	—	—	—	—	42339	8.17	CP	Methyl coenzyme M reductase system	
SSO0170	—	—	—	—	61	25	62804	8.68	AAB	(ilvB-1) Acetolactate synthase large subunit homolog	
SSO0170	—	—	69	28	—	—	62804	8.68	AAB	(ilvB-1) Acetolactate synthase large subunit homolog	
SSO0173	10	23	—	—	44	21	48769	7.23	TRL	(aspS) Aspartyl-tRNA synthetase	
SSO0176	145	31	150	45	—	—	85807	5.75	CP	AAA family ATPase	
SSO0176	68	20	52	16	—	—	85807	5.75	CP	AAA family ATPase	
SSO0191	35	31	—	—	—	—	26715	6.04	HP	Conserved hypothetical protein	
SSO0199	15/150	39/54	—	—	—	—	45410	6.16	AAB	(maT) S-adenosylmethionine synthetase	
SSO0202	41	34	—	—	—	—	24803	5.98	EM	(hpS-2) D-arabino 3-hexulose 6-phosphate formaldehyde lyase	
SSO0205	61	34	—	—	—	—	16694	6.54	HP	Conserved hypothetical protein	
SSO0207	17/96	32/39	147	49	128	47	50264	6.66	CEM	(pmM) Phosphomannomutase	
SSO0215	27	30	—	—	—	—	12112	10.87	TRL	(rps10AB) SSU ribosomal protein S10AB	
SSO0216	20	48	—	—	164	43	48489	9.63	TRL	(tuF-1) Elongation factor 1-alpha (elongation factor tu) (EF-tu)	
SSO0216	—	—	—	—	47	26	48489	9.63	TRL	(tuF-1) Elongation factor 1-alpha (elongation factor tu) (EF-tu)	
SSO0220	31	34	—	—	—	—	13812	10.61	TrR	(nusA) Transcription termination factor nusA homolog, putative	
SSO0223	46	16	—	—	—	—	43805	5.52	TrR	(rpoA2) DNA-directed RNA polymerase, subunit A"	
SSO0225	49	16	—	—	—	—	99674	8.12	TrR	(rpoA1) DNA-directed RNA polymerase, subunit A'	
SSO0228	—	—	—	—	79	33	66933	9.48	TRL	(infB) Translation initiation factor IF2 homolog	
SSO0230	—	—	—	—	48	26	15554	8.40	CIM	(ndk) Nucleoside diphosphate kinase (NDP kinase)	
SSO0230	—	—	—	—	51	29	15554	8.40	CIM	(ndk) Nucleoside diphosphate kinase (NDP kinase)	

Table 12, continued part 2

P2 Tris-Soluble and Tris-Insoluble Proteins Identified by *S. solfataricus* ORF Number, SSO

SSO# Column						IPG Column					
Black SSO# found only in the tris-soluble fraction						Black = ProteinProbe number of peptides found matching					
Red SSO # found only in tris-insoluble fraction						Red = Mascot Mowse matching score, 90% confidence or greater					
Blue SSO# found in both tris-soluble and tris-insoluble fractions						Blue = Mascot Mowse matching score, 95% confidence or greater					
Pct. Column = percentage of the protein identified with MS											
Type = protein functional category											
AAB: Amino Acid Biosynthesis						IS: Insertion Element					
CEM: Cell Envelope Membrane						Nuc: Nucleotide Synthesis/Metabolism					
CIM: Central Intermediary Metabolism						PPM: Proteases and Protein Modifications					
CF: Cofactor Biosynthesis						TrL: Translation					
CP: Cellular Processes						TrR: Transcription and Regulation					
EM: Energy Metabolism						TrSP: Transport					
LM: Lipid Metabolism						RR: Replication and Repair					
HP Hypothetical Protein						Unc: Uncategorized					
S.sol#	IPG 3-10	Pct	IPG 4-7	Pct	IPG 6-11	Pct	MW	pI	Type	Function	
SSO0257	-	-	-	-	45	13	45167	7.56	RR	(cdc6-1)	Cell division control 6/orc1 protein homolog
SSO0278	12	76	102	62	-	-	21376	5.60	PPM	Proteasome subunit	
SSO0278	-	-	35	19	-	-	21376	5.43	PPM	Proteasome subunit	
SSO0279	-	-	105	33	-	-	49144	6.41	TRL	(hisS)	Histidyl-tRNA synthetase
SSO0282	6/117	16/45	91	25	-	-	60367	5.51	TRL	(thsB)	Thermosome beta subunit(thermophilic factor 55)
SSO0286	18	56	-	-	70	37	42708	6.12	HP	Conserved hypothetical protein	
SSO0291	52	57	-	-	-	-	13205	9.07	TrR	(rpoM-1)	DNA-directed RNA polymerase, subunit M
SSO0307	-	-	45	15	-	-	43178	8.34	AAB	(aroC)	Chorismate synthase
SSO0316	8	33	-	-	62	38	24243	7.32	CP	(sod)	Superoxide dismutase [Fe]
SSO0316	-	-	-	-	47	19	24243	7.32	CP	(sod)	Superoxide dismutase [Fe]
SSO0342	-	-	62	46	-	-	11291	4.66	TRL	(rpl12AB)	LSU ribosomal protein L12AB
SSO0344	-	-	-	-	83	48	37061	10.02	TRL	(rplp0)	LSU acidic ribosomal protein P0 homolog
SSO0344	-	-	-	-	88	50	37061	10.02	TRL	(rplp0)	LSU acidic ribosomal protein P0 homolog
SSO0354	13	32	-	-	-	-	43557	6.53	CF	Nicotinic acid phosphoribosyltransferase, putative	
SSO0372	7	63	-	-	-	-	18516	6.64	TrSP	Ferripyochelin binding protein	
SSO0375	-	-	43	18	-	-	33539	10.86	RR	(xerC/D)	XerC/D integrase-recombinase protein
SSO0378	-	-	-	-	117	44	34089	7.91	AAB	(ocd-like1)	Ornithine cyclodeaminase related protein
SSO0397	11	66	-	-	-	-	27535	5.17	RR	(pcnA-like)	Proliferating cell nuclear antigen putative homolog
SSO0407	26/48	60/38	-	-	49	19	38250	8.55	AAB	5-methyltetrahydropteroyltriglu-homocys methyltransferase	
SSO0412	-	-	-	-	47	14	45792	8.48	TRL	(eif2G)	Translation initiation factor eIF 2 gamma
SSO0417	-	-	-	-	59	44	141667	7.42	CP	Phosphonopyruvate decarboxylase	
SSO0421	33/191	52/47	202	38	-	-	86346	6.73	CP	AAA family ATPase	
SSO0421	235	40	181	40	-	-	86346	6.73	CP	AAA family ATPase	
SSO0428	41	45	-	-	-	-	13123	7.04	HP	Hypothetical protein	
SSO0428	69	34	81	58	-	-	13123	7.04	HP	Hypothetical protein	
SSO0452	17	37	73	26	-	-	45448	6.63	TRL	(trpS)	Tryptophanyl-tRNA synthetase
SSO0483	-	-	49	29	-	-	11993	10.15	HP	Hypothetical protein	
SSO0506	-	-	54	40	-	-	38378	8.82	HP	Conserved hypothetical protein	
SSO0527	12	39	97	36	-	-	45166	6.54	EM	Phosphoglycerate kinase	
SSO0530	15/116	43/43	118	35	111	48	48508	6.67	AAB	(glyA)	Serine hydroxymethyltransferase
SSO0555	-	-	34	33	-	-	11707	5.08	HP	Conserved hypothetical protein	
SSO0561	38	32	-	-	-	-	22633	7.32	EM	(atpE)	ATP synthase subunit E
SSO0563	16	40	-	-	-	-	66411	6.30	EM	(atpA)	ATP synthase subunit A
SSO0563	-	-	69	29	71	28	66411	6.30	EM	(atpA)	ATP synthase subunit A
SSO0576	13	55	-	-	92	32	36811	7.08	AAB	(ilvC-1)	Ketol-acid reductoisomerase
SSO0581	-	-	-	-	57	24	30452	10.20	TrR	GTP binding protein, hypothetical	
SSO0589	-	-	43	10	-	-	109103	7.52	TRL	(leuS-2)	Leucyl-tRNA synthetase
SSO0605	48	43	-	-	-	-	13631	4.86	Unc	Zn finger protein, hypothetical	
SSO0606	32	36	-	-	-	-	17975	9.03	TrR	Transcriptional regulator, putative	
SSO0606	-	-	-	-	119	52	17975	9.03	TrR	Transcriptional regulator, putative	
SSO0614	-	-	49	28	-	-	34738	6.54	Nuc	(pyrB)	Asp carbamoyltransferase catalytic chain (ATCase)
SSO0625	-	-	48	17	-	-	38193	10.15	Unc	GTP-binding protein	

Table 12, continued part 3

P2 Tris-Soluble and Tris-Insoluble Proteins Identified by *S. solfataricus* ORF Number, SSO

SSO# Column						IPG Column				
Black SSO# found only in the tris-soluble fraction						Black = ProteinProbe number of peptides found matching				
Red SSO # found only in tris-insoluble fraction						Red = Mascot Mowse matching score, 90% confidence or greater				
Blue SSO# found in both tris-soluble and tris-insoluble fractions						Blue = Mascot Mowse matching score, 95% confidence or greater				
Pct. Column = percentage of the protein identified with MS										
Type = protein functional category										
AAB: Amino Acid Biosynthesis						IS: Insertion Element				
CEM: Cell Envelope Membrane						Nuc: Nucleotide Synthesis/Metabolism				
CIM: Central Intermediary Metabolism						PPM: Proteases and Protein Modifications				
CF: Cofactor Biosynthesis						TrL: Translation				
CP: Cellular Processes						TrR: Transcription and Regulation				
EM: Energy Metabolism						TrSP: Transport				
LM: Lipid Metabolism						RR: Replication and Repair				
HP Hypothetical Protein						Unc: Uncategorized				
S_sol#	IPG 3-10	Pct	IPG 4-7	Pct	IPG 6-11	Pct	MW	pI	Type	Function
SSO0625	-	-	48	17	-	-	38193	10.15	Unc	GTP-binding protein
SSO0654	-	-	44	12	-	-	27980	5.69	LM	(paaF-2) Enoyl CoA hydratase
SSO0694	6	43	-	-	74	48	21326	9.00	Nuc	(adkA) Adenylate kinase
SSO0698	-	-	-	-	65	25	24044	10.24	TRL	(rps5AB) SSU ribosomal protein S5AB
SSO0702	-	-	-	-	85	58	20215	10.39	TRL	(rpl6AB) LSU ribosomal protein L6AB
SSO0713	-	-	-	-	43	38	18126	10.65	TRL	(rpl22AB) LSU ribosomal protein L22AB
SSO0719	-	-	-	-	68	35	39670	10.97	TRL	(rpl3AB) LSU ribosomal protein L3AB
SSO0723	-	-	38	18	-	-	37089	6.34	AAB	(leuB) 3-isopropylmalate dehydrogenase
SSO0728	19	38	119	46	89	38	81820	6.49	TRL	(EF-2) Elongation factor 2
SSO0732	11	41	-	-	-	-	30193	5.01	HP	Conserved hypothetical protein
SSO0737	-	-	-	-	70	30	26199	9.95	HP	Conserved hypothetical protein
SSO0738	11/70	50/58	-	-	-	-	26588	5.82	PPM	Proteasome subunit
SSO0738	98	55	-	-	-	-	26588	5.82	PPM	Proteasome subunit
SSO0747	-	-	40	21	-	-	27289	9.08	HP	Conserved hypothetical protein
SSO0750	-	-	-	-	56	35	24661	8.86	HP	Conserved hypothetical protein
SSO0750	-	-	-	-	41	19	24661	8.86	HP	Conserved hypothetical protein
SSO0755	19/128	58/53	112	43	111	38	48937	7.77	AAB	(ahcY) S-adenosyl-L-homocysteine hydrolase
SSO0758	10/60	41/33	-	-	74	33	26716	7.52	TRL	Peptidyl-prolyl cis-trans isomerase, FKBP-type rotamase
SSO0758	90	22	-	-	78	28	26716	7.52	PPM	Peptidyl-prolyl cis-trans isomerase, FKBP-type rotamase
SSO0766	49	15	-	-	50	33	24903	9.45	PPM	Proteasome subunit
SSO0766	-	-	-	-	55	14	24903	9.45	PPM	Proteasome subunit
SSO0768	-	-	116	42	-	-	37740	5.68	RR	(rfc) Activator 1, replication factor C, small subunit
SSO0770	36	40	-	-	-	-	16838	6.56	CF	(moaC) Molybdenum cofactor biosynthesis protein C
SSO0801	-	-	68	24	-	-	53080	10.63	ISE	Transposase ISC1476
SSO0818	-	-	-	-	57	29	28126	9.12	HP	Conserved hypothetical protein
SSO0838	39	24	-	-	-	-	27177	10.14	HP	Conserved hypothetical protein
SSO0842	-	-	55	22	64	30	53519	10.46	ISE	Transposase ISC1476
SSO0850	-	-	-	-	67	25	35835	10.35	ISE	Transposase ISC1190
SSO0852	44	38	-	-	-	-	17066	10.59	HP	Hypothetical protein
SSO0862	16/126	39/42	121	34	117	34	59674	5.22	TRL	(thsA) Thermosome alpha subunit (thermophilic factor 55)
SSO0862	182	55	47	14	70	16	59674	5.22	TRL	(thsA) Thermosome alpha subunit (thermophilic factor 55)
SSO0873	-	-	-	-	59	23	25050	9.51	HP	Conserved hypothetical protein
SSO0874	-	-	-	-	62	17	38692	7.71	AAB	(asd-1) Aspartate-semialdehyde dehydrogenase
SSO0883	32	46	-	-	-	-	89393	5.87	EM	(ppsA-1) Phosphoenolpyruvate synthase
SSO0883	160	24	101	29	-	-	89393	5.87	EM	(ppsA-1) Phosphoenolpyruvate synthase
SSO0893	-	-	37	9	-	-	47651	5.78	AAB	(trpE) Anthranilate synthase component I
SSO0895	32	31	56	26	-	-	28588	7.14	AAB	(trpC) Indole-3-glycerol phosphate synthase
SSO0907	-	-	-	-	61	31	76573	10.32	RR	(topA) DNA topoisomerase I family A
SSO0907	56	21	-	-	71	22	76573	10.32	RR	(topA) DNA topoisomerase I family A
SSO0909	-	-	60	24	-	-	42248	6.29	CP	AAA family ATPase, p60 katanin
SSO0909	54	23	-	-	-	-	42248	6.29	CP	AAA family ATPase, p60 katanin
SSO0913	12	40	102	55	-	-	46745	5.38	EM	Enolase

Table 12, continued part 4

P2 Tris-Soluble and Tris-Insoluble Proteins Identified by *S. solfataricus* ORF Number, SSO

SSO# Column				IPG Column						
Black SSO# found only in the tris-soluble fraction				Black = ProteinProbe number of peptides found matching						
Red SSO # found only in tris-insoluble fraction				Red = Mascot Mowse matching score, 90% confidence or greater						
Blue SSO# found in both tris-soluble and tris-insoluble fractions				Blue = Mascot Mowse matching score, 95% confidence or greater						
Pct. Column = percentage of the protein identified with MS										
Type = protein functional category										
AAB: Amino Acid Biosynthesis				IS: Insertion Element						
CEM: Cell Envelope Membrane				Nuc: Nucleotide Synthesis/Metabolism						
CIM: Central Intermediary Metabolism				PPM: Proteases and Protein Modifications						
CF: Cofactor Biosynthesis				TrL: Translation						
CP: Cellular Processes				TrR: Transcription and Regulation						
EM: Energy Metabolism				TrSP: Transport						
LM: Lipid Metabolism				RR: Replication and Repair						
HP Hypothetical Protein				Unc: Uncategorized						
S.sol#	IPG 3-10	Pct	IPG 4-7	Pct	IPG 6-11	Pct	MW	pI	Type	Function
SSO0913	12	40	102	55	—	—	46745	5.38	EM	Enolase
SSO0921	9	37	—	—	—	—	28178	6.70	HP	Conserved hypothetical protein
SSO0925	—	—	—	—	60	53	17769	9.01	TrSP	ABC transporter
SSO0925	80	51	—	—	74	44	17769	9.01	TrSP	ABC transporter
SSO0940	51	23	—	—	69	33	26422	8.08	TrR	Fibrillarin-like pre-rRNA processing protein
SSO0940	—	—	—	—	47	25	26422	8.08	TrR	Fibrillarin-like pre-rRNA processing protein
SSO0955	85	59	—	—	69	31	19586	6.30	HP	Hypothetical protein
SSO0955	—	—	—	—	67	54	19586	6.30	HP	Hypothetical protein
SSO0959	42	17	61	23	—	—	64175	10.08	RR	DNA repair protein rad25
SSO0962	—	—	—	—	50	30	10974	11.11	Unc	DNA binding protein SSO10b
SSO0967	—	—	—	—	102	38	35050	7.87	Unc	Deoxyhypusine synthase
SSO0970	6	63	—	—	—	—	14492	6.55	Unc	(eIF5A) Initiation factor 5A, hypothetical
SSO0974	17	47	—	—	52	25	31658	9.59	Unc	3-hydroxyisobutyrate dehydrogenase
SSO0981	—	—	—	—	64	34	49883	10.05	EM	(pyK) Pyruvate kinase
SSO0983	—	—	51	30	—	—	19391	7.55	HP	Hypothetical protein
SSO0988	50	12	—	—	—	—	53641	5.75	EM	Alpha amylase
SSO0989	—	—	—	—	84	24	48225	9.86	CEM	Sugar phosphate nucleotydyl transferase
SSO0999	—	—	40	15	—	—	53282	4.78	TrSP	ABC Transporter
SSO1003	—	—	—	—	140	44	36800	9.54	TrSP	Sugar ABC transporter
SSO10285	9/102	33/64	—	—	—	—	10734	4.23	HP	Hypothetical protein
SSO1045	—	—	—	—	44	13	32536	6.54	CIM	Ribose phosphate pyrophosphokinase
SSO1047	5	23	—	—	—	—	27567	4.36	RR	(pcnA-2) Proliferating cell nuclear antigen homolog (PCNA)
SSO1067	7	62	—	—	—	—	19227	7.55	PPM	Intracellular proteinase
SSO1074	42	37	—	—	—	—	26089	5.10	HP	Conserved hypothetical protein
SSO10802	—	—	—	—	39	45	7525	10.60	Unc	(cutC-2) CODH, small chain. Carboxy-end fragment
SSO1806	25	42	—	—	—	—	72695	6.21	HP	Conserved hypothetical protein
SSO1088	9	40	56	23	—	—	23232	4.74	HP	Conserved hypothetical protein
SSO1095	12/57	22/16	145	31	—	—	95774	6.58	EM	Aconitate hydratase
SSO1098	65	41	60	47	49	26	18157	6.87	HP	Conserved hypothetical protein
SSO1101	5	36	—	—	—	—	14596	8.99	TrR	Transcriptional regulator, putative
SSO1108	5	34	—	—	—	—	14071	8.08	TrR	Transcriptional regulator Lrs14
SSO1119	56	26	48	22	—	—	46460	4.72	HP	Hypothetical protein
SSO1126	59	55	65	46	—	—	16112	5.34	HP	Conserved hypothetical protein
SSO1107	34	29	—	—	—	—	13701	9.65	HP	Hypothetical protein
SSO1144	64	34	—	—	—	—	31650	9.32	TrSP	ABC transporter, ATP binding protein
SSO1144	—	—	—	—	51	17	31650	9.32	TrSP	ABC transporter, ATP binding protein
SSO1145	47	5	—	—	—	—	31650	6.43	AAB	(trpB-like) Tryptophan synthase beta chain, hypothetical
SSO1151	11/54	35/31	79	33	—	—	51710	4.99	HP	Conserved hypothetical protein
SSO1152	—	—	—	—	70	23	46122	9.50	HP	Conserved hypothetical protein
SSO1155	—	—	—	—	51	16	50734	8.67	HP	Conserved hypothetical protein
SSO1157	—	—	—	—	52	19	23712	9.98	HP	Conserved hypothetical protein

Table 12, continued part 5

P2 Tris-Soluble and Tris-Insoluble Proteins Identified by *S. solfataricus* ORF Number, SSO

SSO# Column				IPG Column							
Black SSO# found only in the tris-soluble fraction				Black = ProteinProbe number of peptides found matching							
Red SSO # found only in tris-insoluble fraction				Red = Mascot Mowse matching score, 90% confidence or greater							
Blue SSO# found in both tris-soluble and tris-insoluble fractions				Blue = Mascot Mowse matching score, 95% confidence or greater							
Pct. Column = percentage of the protein identified with MS											
Type = protein functional category											
AAB: Amino Acid Biosynthesis						IS: Insertion Element					
CEM: Cell Envelope Membrane						Nuc: Nucleotide Synthesis/Metabolism					
CIM: Central Intermediary Metabolism						PPM: Proteases and Protein Modifications					
CF: Cofactor Biosynthesis						TrL: Translation					
CP: Cellular Processes						TrR: Transcription and Regulation					
EM: Energy Metabolism						TrSP: Transport					
LM: Lipid Metabolism						RR: Replication and Repair					
HP Hypothetical Protein						Unc: Uncategorized					
S.sol#	IPG 3-10	Pct	IPG 4-7	Pct	IPG 6-11	Pct	MW	pI	Type	Function	
SSO11575	25	28	-	-	-	-	8356	4.81	HP	Hypothetical protein	
SSO1168	60	17	-	-	110	44	41884	9.26	TrSP	Sugar ABC transporter	
SSO1184	11/58	40/52	-	-	121	62	21251	9.64	CF	(entB-like1) Isochorismatase related protein	
SSO1197	-	-	-	-	32	29	17803	10.06	HP	Conserved hypothetical protein	
SSO1208	-	-	27	32	-	-	20965	5.22	EM	(porG-1) Pyruvic-ferredoxin oxidoreductase gamma chain	
SSO1225	7	23	58	26	-	-	45554	8.32	CP	(todA) Toluene 1,2-dioxygenase syst ferredoxin-NAD(+) red.	
SSO1237	-	-	-	-	80	59	17492	10.13	ISE	First ORF in transposon ISC1491	
SSO1254	14	62	47	40	-	-	30620	6.58	EM	(citE) Citryl-CoA lyase beta subunit homolog	
SSO1272	10	27	-	-	-	-	38920	6.10	HP	Conserved hypothetical protein	
SSO1273	-	-	55	8	-	-	97957	5.22	HP	Conserved hypothetical protein	
SSO1322	5	18	-	-	-	-	37185	5.44	AAB	(ilvC-2) Ketol-acid reductoisomerase	
SSO1322	-	-	35	12	-	-	37185	5.44	AAB	(ilvC-2) Ketol-acid reductoisomerase	
SSO1334	-	-	54	14	59	18	94701	7.25	EM	(aceB/mas) Malate synthase, putative	
SSO1347	48	25	-	-	-	-	21819	11.15	ISE	Third ORF in transposon ISC1904	
SSO1355	-	-	91	36	-	-	43068	6.31	PPM	(cpsA-1) Thermostable carboxypeptidase	
SSO1391	-	-	34	16	-	-	33490	6.27	HP	Conserved hypothetical protein	
SSO1423	53	30	-	-	-	-	23104	10.02	HP	Hypothetical protein	
SSO1423	47	32	-	-	-	-	23104	10.02	HP	Hypothetical protein	
SSO1426	5/49	15/22	-	-	-	-	31527	9.68	HP	Hypothetical protein	
SSO1442	13	52	-	-	111	43	35264	6.57	HP	Conserved hypothetical protein	
SSO1445	43	27	-	-	-	-	26816	10.05	HP	Conserved hypothetical protein	
SSO1448	56	21	-	-	-	-	45114	10.58	ISE	Transposase ISC1359	
SSO1468	-	-	75	30	-	-	48256	9.97	HP	Conserved hypothetical protein	
SSO1468	72	22	-	-	-	-	48256	9.97	HP	Conserved hypothetical protein	
SSO1500	58	14	-	-	-	-	121122	9.10	HP	Hypothetical protein	
SSO1512	-	-	-	-	63	24	54120	8.00	HP	Hypothetical protein	
SSO1513	10	51	-	-	-	-	17457	9.55	HP	Hypothetical protein	
SSO1513	-	-	-	-	72	36	17457	9.55	HP	Hypothetical protein	
SSO1514	37	18	-	-	-	-	29872	7.35	HP	Conserved hypothetical protein	
SSO1514	7/55	34/49	76	41	88	42	29872	7.35	HP	Conserved hypothetical protein	
SSO1526	-	-	77	28	-	-	35323	4.99	EM	(pdhB-2) Pyruvate dehydrogenase, beta subunit (lipoamide).	
SSO1532	15/47	65/40	44	27	53	20	39215	6.82	HP	Conserved hypothetical protein	
SSO1533	19	47	56	18	-	-	44561	6.72	Unc	Homogentisate oxygenase related protein	
SSO1533	-	-	45	15	-	-	44561	6.72	Unc	Homogentisate oxygenase related protein	
SSO1553	57	45	-	-	-	-	20356	10.19	HP	Hypothetical protein	
SSO1553	43	15	-	-	-	-	20356	10.19	HP	Hypothetical protein	
SSO1556	64	40	92	51	72	45	28343	6.10	ISE	First ORF in transposon ISC1491	
SSO1629	12	35	-	-	-	-	51529	8.91	EM	(gapN-1) Glyceraldehyde-3-phos. dehydrogenase, NADP dep.	
SSO1662	12/81	28/32	62	20	-	-	63396	5.28	AAB	(huyB-1) N-methylhydantoinase B	
SSO1663	27	43	49	25	-	-	76406	5.76	AAB	(huyA-1) N-methylhydantoinase A	
SSO1668	74	38	-	-	-	-	57620	6.83	AAB	(huyA-like) N-methylhydantoinase A related protein	

Table 12, continued part 6

P2 Tris-Soluble and Tris-Insoluble Proteins Identified by *S. solfataricus* ORF Number, SSO

SSO# Column			IPG Column								
Black SSO# found only in the tris-soluble fraction			Black = ProteinProbe number of peptides found matching								
Red SSO # found only in tris-insoluble fraction			Red = Mascot Mowse matching score, 90% confidence or greater								
Blue SSO# found in both tris-soluble and tris-insoluble fractions			Blue = Mascot Mowse matching score, 95% confidence or greater								
Pct. Column = percentage of the protein identified with MS											
Type = protein functional category											
AAB: Amino Acid Biosynthesis						IS: Insertion Element					
CEM: Cell Envelope Membrane						Nuc: Nucleotide Synthesis/Metabolism					
CIM: Central Intermediary Metabolism						PPM: Proteases and Protein Modifications					
CF: Cofactor Biosynthesis						TrL: Translation					
CP: Cellular Processes						TrR: Transcription and Regulation					
EM: Energy Metabolism						TrSP: Transport					
LM: Lipid Metabolism						RR: Replication and Repair					
HP Hypothetical Protein						Unc: Uncategorized					
S_sol#	IPG 3-10	Pct	IPG 4-7	Pct	IPG 6-11	Pct	MW	pI	Type	Function	
SSO1680	38	24	—	—	72	66	11486	10.09	HP	Conserved hypothetical protein	
SSO1704	—	—	52	27	—	—	28270	7.43	ISE	First ORF in transposon ISC1229	
SSO1712	20	46	53	19	—	—	52054	7.58	HP	Hypothetical protein	
SSO1772	65	48	—	—	—	—	28361	6.10	ISE	First ORF in transposon ISC1491	
SSO1811	13	33	84	34	—	—	44087	5.77	PPM	(pepQ-like1) Prolidase (Xaa-Pro dipeptidase)	
SSO1817	96	16	—	—	—	—	33655	6.90	CIM	(cysA-2) Thiosulfate sulfurtransferase	
SSO1824	52	19	—	—	68	30	49215	6.69	HP	Conserved hypothetical protein	
SSO1824	—	—	27	7	—	—	49215	6.69	HP	Conserved hypothetical protein	
SSO1829	50	19	44	15	—	—	50004	10.90	ISE	Second ORF in transposon ISC1904	
SSO1842	12	39	45	10	53	31	15226	9.02	EM	(gapN-2) Glyceraldehyde-3-phos. dehydrogenase, NADP dep	
SSO1865	9/79	27/37	—	—	80	37	15226	9.02	HP	Conserved hypothetical protein	
SSO1865	69	30	—	—	39	30	15226	9.02	HP	Conserved hypothetical protein	
SSO1889	82	35	54	42	—	—	38702	10.09	Unc	ATP-dependent RNA helicase	
SSO1907	19/100	45/40	92	32	—	—	46031	6.98	AAB	(gdhA-2) NAD specific glutamate dehydrogenase	
SSO1907	105	31	89	40	70	27	46031	6.98	AAB	(gdhA-2) NAD specific glutamate dehydrogenase	
SSO1908	46	59	—	—	—	—	11475	9.00	HP	Hypothetical protein	
SSO1917	39	28	—	—	—	—	12489	9.85	HP	Hypothetical protein	
SSO1919	24	34	—	—	—	—	12777	10.83	ISE	First ORF in transposon ISC1395	
SSO1947	39	22	34	21	—	—	17105	6.57	HP	Hypothetical protein	
SSO1960	6	36	66	44	51	31	31706	6.82	HP	Hypothetical protein	
SSO1960	84	36	47	25	—	—	31706	6.82	HP	Hypothetical protein	
SSO2017	18/93	55/49	45	28	—	—	19904	8.40	LM	(paaf-4) Enoyl CoA hydratase	
SSO2044	17/48	50/33	100	28	75	32	46063	6.98	AAB	(gdhA-4) NAD specific glutamate dehydrogenase	
SSO2044	99	33	56	22	84	27	46063	6.98	AAB	(gdhA-4) NAD specific glutamate dehydrogenase	
SSO2059	16	34	—	—	—	—	70115	8.10	LM	(acsA-6) Acetyl-CoA synthetase (acetate-CoA ligase)	
SSO2060	—	—	49	15	—	—	41394	5.98	LM	(acd-3) Acyl-CoA dehydrogenase	
SSO2061	—	—	110	27	—	—	41597	5.39	LM	(acaB-2) Acetyl-CoA C-acetyltransferase	
SSO2062	—	—	—	—	149	38	44865	7.70	LM	(acaB-3) Acetyl-CoA C-acetyltransferase	
SSO2063	8/54	16/18	—	—	—	—	19095	8.96	HP	Conserved hypothetical protein	
SSO2091	—	—	37	18	—	—	76017	5.57	HP	(cutA-2) Carbon monoxide dehydrogenase, large chain	
SSO2095	—	—	64	28	—	—	86609	6.39	EM	(treY) Maltotriose synthase	
SSO2102	—	—	48	27	—	—	21678	10.39	HP	Hypothetical protein	
SSO2112	—	—	—	—	53	28	31017	8.88	HP	Conserved hypothetical protein	
SSO2121	11	45	—	—	—	—	24745	7.55	CP	(bcp-2) Peroxiredoxin, bacterioferritin comigratory protein	
SSO2122	—	—	57	13	—	—	55655	6.23	TRL	(gatA-3) Glutamyl-tRNA amidotransferase, subunit A	
SSO2141	6	15	125	36	—	—	64282	6.30	PPM	(apeH-2) Acylaminoacyl-peptidase, putative	
SSO2150	13	24	99	28	—	—	80646	6.05	EM	(cutA-3) Carbon monoxide dehydrogenase, large chain	
SSO2154	24/51	37/40	188	32	—	—	90515	5.64	PPM	Tricorn protease interacting factor F2	
SSO2190	10	43	—	—	—	—	27391	6.52	HP	Hypothetical protein	
SSO2195	—	—	63	34	116	47	29663	9.91	CEM	Erythrocyte band 7 membrane protein homolog	
SSO2199	23	33	—	—	—	—	118727	5.35	HP	Conserved hypothetical protein	

Table 12, continued part 7

P2 Tris-Soluble and Tris-Insoluble Proteins Identified by *S. solfataricus* ORF Number, SSO

SSO# Column				IPG Column							
Black SSO# found only in the tris-soluble fraction				Black = ProteinProbe number of peptides found matching							
Red SSO # found only in tris-insoluble fraction				Red = Mascot Mowse matching score, 90% confidence or greater							
Blue SSO# found in both tris-soluble and tris-insoluble fractions				Blue = Mascot Mowse matching score, 95% confidence or greater							
Pct. Column = percentage of the protein identified with MS											
Type = protein functional category											
AAB: Amino Acid Biosynthesis						IS: Insertion Element					
CEM: Cell Envelope Membrane						Nuc: Nucleotide Synthesis/Metabolism					
CIM: Central Intermediary Metabolism						PPM: Proteases and Protein Modifications					
CF: Cofactor Biosynthesis						TrL: Translation					
CP: Cellular Processes						TrR: Transcription and Regulation					
EM: Energy Metabolism						TrSP: Transport					
LM: Lipid Metabolism						RR: Replication and Repair					
HP Hypothetical Protein						Unc: Uncategorized					
S.soi#	IPG 3-10	Pct	IPG 4-7	Pct	IPG 6-11	Pct	MW	pl	Type	Function	
SSO2205	5/62	25/38	—	—	—	—	25945	6.53	LM	(fabG-4) 3-oxoacyl-(acyl carrier protein) reductase	
SSO2208	—	—	50	33	—	—	27264	9.43	HP	Conserved hypothetical protein	
SSO2222	12	47	—	—	89	31	36542	9.28	Nuc	(trxB-1) Thioredoxin reductase	
SSO2241	8/122	24/38	114	28	—	—	68436	5.41	Unc	(bps2) BPS2 protein homolog	
SSO2241	173	40	48	12	54	11	68436	5.41	Unc	(bps2) BPS2 protein homolog	
SSO2249	69	18	—	—	60	17	101601	6.19	RR	Purine NTPase	
SSO2254	—	—	55	36	—	—	54308	6.55	HP	Conserved hypothetical protein	
SSO2276	—	—	—	—	95	34	28059	9.77	LM	(fabG-5) 3-oxoacyl-(acyl carrier protein) reductase	
SSO2281	—	—	—	—	75	39	33141	7.82	HP	Conserved hypothetical protein	
SSO2282	5	43	—	—	—	—	16310	5.35	HP	Hypothetical protein	
SSO2301	—	—	48	25	—	—	24701	7.73	CF	(cblL) Cobalamin biosynth precomin-2 methyltransferase	
SSO2345	52	28	—	—	—	—	32274	5.16	LM	(gdS-2) Geranylgeranyl pyrophosphate synthetase	
SSO2345	—	—	40	15	—	—	32274	5.16	LM	(gdS-2) Geranylgeranyl pyrophosphate synthetase	
SSO2347	—	—	—	—	46	42	17786	9.46	TrR	(asnC) Transcriptional regulatory protein, AsnC family	
SSO2353	42	21	—	—	—	—	44951	7.50	LM	Geranylgeranyl hydrogenase, putative	
SSO2359	52	42	—	—	—	—	14236	6.04	EM	(sdhD) Succinate dehydrogenase subunit D	
SSO2363	150	43	106	44	—	—	43683	6.18	HP	Conserved hypothetical protein	
SSO2364	6	53	—	—	—	—	16138	9.82	RR	(SSB) Single-stranded DNA binding protein	
SSO2390	6/55	35/44	70	35	—	—	19694	4.78	EM	(ppa) Inorganic pyrophosphatase, putative	
SSO2401	—	—	47	21	—	—	26703	5.89	HP	Conserved hypothetical protein	
SSO2416	8	35	—	—	—	—	35221	9.05	Nuc	(trxB-3) Thioredoxin reductase	
SSO2420	36/61	40/20	48	11	88	50	68396	9.19	CP	AAA family ATPase	
SSO2425	—	—	65	32	—	—	64121	6.32	CP	(mcmA1) Methylmalonyl-CoA mutase, alpha-subunit, chain /	
SSO2427	118	47	51	32	—	—	20097	5.10	CP	Small heat shock protein hsp20 family	
SSO2432	74	49	—	—	—	—	21376	7.91	HP	Conserved hypothetical protein	
SSO2434	11	33	—	—	89	54	30743	10.04	EM	(cutB-1) Carbon monoxide dehydrogenase, medium chain.	
SSO2450	19/161	50/52	187	56	—	—	53201	8.11	TrR	TATA binding protein (TBP)-interacting protein (TIP49-like)	
SSO2450	86	38	—	—	94	40	53201	8.11	TrR	TATA binding protein (TBP)-interacting protein (TIP49-like)	
SSO2483	—	—	60	26	50	30	37354	5.50	EM	(sucC) Succinyl-CoA synthetase, beta subunit	
SSO2500	9/76	40/41	—	—	51	31	26848	8.59	LM	(fabG-7) 3-oxoacyl-(acyl carrier protein) reductase	
SSO2509	30	33	—	—	—	—	14786	8.85	HP	Hypothetical protein	
SSO2510	—	—	52	10	—	—	62390	6.93	LM	(alkK-2) Medium-chain-fatty-acid-CoA ligase	
SSO2514	24/117	43/37	—	—	159	36	73038	8.29	LM	3-hydroxyacyl-CoA dehydrogenase/enoyl CoA hydratase	
SSO2522	—	—	—	—	105	29	35565	6.53	CP	(php) Aryldialkylphosphatase (phosphotriesterase)	
SSO2522	43	20	—	—	—	—	35565	6.53	CP	(php) Aryldialkylphosphatase (phosphotriesterase)	
SSO2537	59	24	—	—	—	—	69056	6.77	EM	Phosphoenolpyruvate carboxykinase (GTP)	
SSO2583	—	—	45	14	—	—	44686	6.76	LM	(sqdB) Sulfolipid biosynthesis protein	
SSO2585	12/99	44/40	—	—	139	42	36394	8.86	EM	(sqdB) L-lactate dehydrogenase	
SSO2586	14	38	153	60	78	36	48672	6.56	HP	Conserved hypothetical protein	
SSO2589	8	27	86	27	—	—	42752	7.86	EM	Citrate synthase	
SSO2597	13	43	—	—	139	40	41976	7.91	AAB	(agxT) Serine-pyruvate aminotransferase	

Table 12, continued part 8

P2 Tris-Soluble and Tris-Insoluble Proteins Identified by *S. solfataricus* ORF Number, SSO

SSO# Column						IPG Column					
Black SSO# found only in the tris-soluble fraction						Black = ProteinProbe number of peptides found matching					
Red SSO # found only in tris-insoluble fraction						Red = Mascot Mowse matching score, 90% confidence or greater					
Blue SSO# found in both tris-soluble and tris-insoluble fractions						Blue = Mascot Mowse matching score, 95% confidence or greater					
Pct. Column = percentage of the protein identified with MS											
Type = protein functional category											
AAB: Amino Acid Biosynthesis						IS: Insertion Element					
CEM: Cell Envelope Membrane						Nuc: Nucleotide Synthesis/Metabolism					
CIM: Central Intermediary Metabolism						PPM: Proteases and Protein Modifications					
CF: Cofactor Biosynthesis						TrL: Translation					
CP: Cellular Processes						TrR: Transcription and Regulation					
EM: Energy Metabolism						TrSP: Transport					
LM: Lipid Metabolism						RR: Replication and Repair					
HP Hypothetical Protein						Unc: Uncategorized					
S_sol#	IPG 3-10	Pct	IPG 4-7	Pct	IPG 6-11	Pct	MW	pI	Type	Function	
SSO2603	5	31	--	--	--	--	14144	9.05	CP	Small heat shock protein hsp20 family	
SSO2613	10	59	49	17	51	33	17461	8.05	CP	(bcp-4) Peroxiredoxin, bacterioferitin comigratory homolog	
SSO2613	41	26	--	--	--	--	17461	8.05	CP	(bcp-4) Peroxiredoxin, bacterioferitin comigratory homolog	
SSO2623	--	--	65	40	--	--	28058	5.69	LM	(paaf-6) Enoyl CoA hydratase	
SSO2625	--	--	--	--	82	23	41892	7.26	LM	(acaB-7) Acetyl-CoA c-acetyltransferase	
SSO2626	--	--	--	--	56	45	14574	8.49	HP	Conserved hypothetical protein	
SSO2628	--	--	63	50	--	--	13453	6.23	HP	Hypothetical protein	
SSO2629	12	40	--	--	76	51	43054	7.16	EM	Oxidoreductase (flavoprotein)	
SSO2629	62	36	--	--	46	16	43054	7.16	EM	Oxidoreductase (flavoprotein)	
SSO2632	6	19	--	--	--	--	42230	4.75	HP	Hypothetical protein	
SSO2636	--	--	--	--	139	45	31051	10.14	EM	(cutB-2) Carbon monoxide dehydrogenase, medium chain.	
SSO2639	21/118	40/33	153	40	122	40	83013	6.35	EM	(cutA-4) Carbon monoxide dehydrogenase, large chain	
SSO2642	56	50	57	44	--	--	16034	5.43	EM	(nr) Rubrerythrin	
SSO2644	11/73	54/48	97	60	--	--	22420	5.53	HP	Hypothetical protein	
SSO2645	8/49	26/30	--	--	--	--	21739	10.43	HP	Hypothetical protein	
SSO2647	35	13	--	--	--	--	34906	9.29	TrSP	ABC transporter, ATP binding protein	
SSO2651	9	11	--	--	--	--	83007	6.66	TrSP	(pacS) Cation transporting ATPase	
SSO2675	--	--	104	35	--	--	90142	5.56	PPM	Tricorn protease interacting factor F3	
SSO2677	--	--	--	--	48	26	18994	9.79	HP	Conserved hypothetical protein	
SSO2693	--	--	50	21	--	--	67147	6.30	PPM	(apeH-3) Acylaminoacyl-peptidase, putative	
SSO2740	12	63	--	--	--	--	27595	8.92	HP	Conserved hypothetical protein	
SSO2760	--	--	82	28	--	--	84891	5.97	EM	(cutA-5) Carbon monoxide dehydrogenase, large chain	
SSO2761	24	67	--	--	--	--	44363	6.64	LM	(acd-5) Acyl-CoA dehydrogenase	
SSO2762	10	41	--	--	--	--	30620	10.15	EM	(etfA) Electron transfer flavoprotein, subunit alpha	
SSO2763	10/73	39/47	--	--	--	--	26560	7.08	EM	(etfB) Electron transfer flavoprotein, subunit beta	
SSO2773	--	--	72	34	--	--	34592	5.94	CIM	(hpcE-2) 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase	
SSO2776	61	19	--	--	--	--	45243	5.69	EM	Electron transfer oxidoreductase	
SSO2777	--	--	46	21	--	--	13656	5.10	HP	Hypothetical protein	
SSO2800	--	--	--	--	48	20	36851	8.32	EM	(adh-12) Alcohol dehydrogenase (Zn containing)	
SSO2813	--	--	51	37	--	--	18603	6.54	HP	Conserved hypothetical protein	
SSO2815	16/61	35/26	--	--	116	44	70228	6.82	EM	2-oxoacid-ferredoxin oxidoreductase, alpha chain	
SSO2817	--	--	--	--	91	23	67141	6.31	EM	(etfAB/fixAB) Electron transfer flavoprotein alpha/ beta-subur	
SSO2819	--	--	--	--	51	34	43305	9.64	EM	(fixC) FIXC protein homolog	
SSO2819	--	--	--	--	127	46	43305	9.64	EM	(fixC) FIXC protein homolog	
SSO2820	--	--	56	26	--	--	34975	5.44	EM	(ppsA-2) Phosphoenolpyruvate synthase	
SSO2827	--	--	40	45	--	--	15375	9.85	HP	Conserved hypothetical protein	
SSO2827	50	46	--	--	41	31	15375	9.85	HP	Conserved hypothetical protein	
SSO2831	--	--	--	--	72	19	66977	8.69	CP	AAA family ATPase	
SSO2850	--	--	--	--	54	21	39113	8.85	TrSP	ABC transporter, ATP binding protein (glucose)	
SSO2850	83	33	--	--	116	41	39113	8.85	TrSP	ABC transporter, ATP binding protein (glucose)	
SSO2863	28/53	41/18	110	47	--	--	75799	6.89	LM	(acsA-9) Acetyl-CoA synthetase (acetate-CoA ligase)	

Table 12, continued part 9

P2 Tris-Soluble and Tris-Insoluble Proteins Identified by *S. solfataricus* ORF Number, SSO

SSO# Column						IPG Column					
Black SSO# found only in the tris-soluble fraction						Black = ProteinProbe number of peptides found matching					
Red SSO # found only in tris-insoluble fraction						Red = Mascot Mowse matching score, 90% confidence or greater					
Blue SSO# found in both tris-soluble and tris-insoluble fractions						Blue = Mascot Mowse matching score, 95% confidence or greater					
Pct. Column = percentage of the protein identified with MS											
Type = protein functional category											
AAB: Amino Acid Biosynthesis						IS: Insertion Element					
CEM: Cell Envelope Membrane						Nuc: Nucleotide Synthesis/Metabolism					
CIM: Central Intermediary Metabolism						PPM: Proteases and Protein Modifications					
CF: Cofactor Biosynthesis						TrL: Translation					
CP: Cellular Processes						TrR: Transcription and Regulation					
EM: Energy Metabolism						TrSP: Transport					
LM: Lipid Metabolism						RR: Replication and Repair					
HP Hypothetical Protein						Unc: Uncategorized					
S.sol#	IPG 3-10	Pct	IPG 4-7	Pct	IPG 6-11	Pct	MW	pI	Type	Function	
SSO2865	-	-	55	44	-	-	12021	9.07	HP	Hypothetical protein	
SSO2867	-	-	-	-	41	27	13365	7.30	HP	Hypothetical protein	
SSO2893	-	-	54	38	-	-	11470	10.84	ISE	Second ORF in transposon ISC1190	
SSO2909	-	-	-	-	42	12	50601	6.79	CIM	(cysI) Sulfite reductase hemoprotein beta component	
SSO2918	51	45	-	-	-	-	14987	10.21	ISE	First ORF in transposon ISC1212	
SSO2963	-	-	31	38	-	-	22481	8.38	HP	Hypothetical protein	
SSO2976	46	24	-	-	-	-	15827	5.47	HP	Hypothetical protein	
SSO2978	64	29	-	-	-	-	40339	6.83	TRL	(eIF2B) Translation initiation factor aIF-2B, subunit 1	
SSO2981	48	19	-	-	-	-	35164	9.85	EM	Epimerase, putative	
SSO2996	17	41	-	-	-	-	43166	5.93	LM	(hdb-3) 3-hydroxyacyl-CoA dehydrogenase	
SSO3006	10	12	133	29	-	-	111159	6.12	CIM	Alpha-mannosidase	
SSO3009	47	17	70	21	88	21	79542	6.41	EM	(cutA-7) Carbon monoxide dehydrogenase, large chain	
SSO3019	14/47	28/22	48	16	-	-	56691	6.47	CIM	(lacS) Beta-glycosidase	
SSO3026	34	42	-	-	-	-	12596	7.78	ISE	One of two inversly orientated ORFs in ISC1043	
SSO3032	-	-	37	21	-	-	83618	6.13	CIM	Beta-xylosidase	
SSO3051	16	27	-	-	-	-	80441	5.61	CIM	(malA) Alpha-glucosidase	
SSO3080	-	-	-	-	52	32	25793	9.66	HP	Conserved hypothetical protein	
SSO3105	85	33	-	-	119	31	54841	6.60	PPM	Thermostable carboxypeptidase 1	
SSO3107	44	17	-	-	-	-	59421	7.36	AAB	(ilvD) Dihydroxy-acid dehydratase	
SSO3153	-	-	-	-	67	26	49977	10.83	ISE	Second ORF in transposon ISC1904	
SSO3161	37	13	-	-	-	-	38749	9	EM	(glcF) Glycolate oxidase iron-sulfur subunit.	
SSO3169	35	13	-	-	-	-	30606	8.86	TrSP	ABC transporter, ATP binding protein related protein	
SSO3194	-	-	97	44	-	-	56926	6.16	EM	(gapN-3) Glyceraldehyde-3-phos. dehydrogenase, NADP de	
SSO3195	-	-	126	35	-	-	34874	6.10	EM	Fructokinase	
SSO3197	-	-	-	-	68	35	34795	7.03	EM	(eda) 2-keto-3-deoxy gluconate aldolase	
SSO3198	17	44	-	-	-	-	44729	5.36	CIM	Muconate cycloisomerase related protein	
SSO3205	-	-	-	-	60	35	31257	7.44	HP	Conserved hypothetical protein	
SSO3206	5/52	53/60	-	-	-	-	14249	6.95	TRL	Protein synthesis inhibitor, putative	
SSO3210	-	-	101	24	-	-	61832	6.19	AAB	(ilvB-6) Acetolactate synthase large subunit homolog	
SSO3211	18	42	60	21	-	-	46521	6.16	AAB	(gabT-2) 4-aminobutyrate aminotransferase	
SSO3214	-	-	50	9	-	-	87982	7.22	HP	Conserved hypothetical protein	
SSO3216	-	-	-	-	72	25	53111	5.80	CF	Gamma-glutamyltranspeptidase, hypothetical	
SSO3226	-	-	-	-	107	27	29915	8.84	HP	Conserved hypothetical protein	
SSO3239	-	-	76	22	-	-	79023	6.12	EM	(cutA-8) Carbon monoxide dehydrogenase, large chain	
SSO5468	-	-	-	-	28	42	9670	10.69	TrR	(rpoH) DNA-directed RNA polymerase, subunit H	
SSO5522	5	72	40	20	56	83	9396	6.91	TrR	Regulatory protein, AsnC family, putative	
SSO5576	-	-	-	-	39	41	10446	8.63	TrR	(rpoM-like) DNA-directed RNA polymerase, putative subunit	
SSO5668	-	-	-	-	87	61	10606	10.41	TRL	(rplX) LSU ribosomal protein LX	
SSO6877	-	-	-	-	47	28	10233	9.39	Unc	DNA binding protein SSO10b	
SSO6877	-	-	-	-	54	39	10233	9.39	Unc	DNA binding protein SSO10b	
SSO8813	38	62	-	-	-	-	10475	9.98	HP	Conserved hypothetical protein	

Table 12, continued part 10

P2 Tris-Soluble and Tris-Insoluble Proteins Identified by *S. solfataricus* ORF Number, SSO**SSO# Column**

Black SSO# found only in the tris-soluble fraction

Red SSO # found only in tris-insoluble fraction

Blue SSO# found in both tris-soluble and tris-insoluble fractions

IPG Column

Black = ProteinProbe number of peptides found matching

Red = Mascot Mowse matching score, 90% confidence or greater

Blue = Mascot Mowse matching score, 95% confidence or greater

Pct. Column = percentage of the protein identified with MS**Type = protein functional category**

AAB: Amino Acid Biosynthesis

CEM: Cell Envelope Membrane

CIM: Central Intermediary Metabolism

CF: Cofactor Biosynthesis

CP: Cellular Processes

EM: Energy Metabolism

LM: Lipid Metabolism

HP Hypothetical Protein

IS: Insertion Element

Nuc: Nucleotide Synthesis/Metabolism

PPM: Proteases and Protein Modifications

TrL: Translation

TrR: Transcription and Regulation

TrSP: Transport

RR: Replication and Repair

Unc: Uncategorized

S.sol#	IPG 3-10	Pct	IPG 4-7	Pct	IPG 6-11	Pct	MW	pI	Type	Function
SSO8936	-	-	72	79	-	-	9569	9.12	ISE	First ORF in transposon ISC1212
SSO8936	-	-	49	62	-	-	9569	9.12	ISE	First ORF in transposon ISC1212
SSO8938	37	34	-	-	-	-	9053	6.79	HP	Hypothetical protein
SSO9180	-	-	-	-	43	46	7279	10.59	Unc	(ssh7A /Sso7d-2) 7 KD DNA-binding protein (SSO7D)
SSO9268	31	37	-	-	-	-	10627	10.11	HP	Hypothetical protein

CHAPTER 6

COMPARISONS BETWEEN SULFOLOBUS SOLFATARICUS
STRAINS P1 AND P2 USING 2-DE AND MASS SPECTROMETRYChapter Summary

We have shown that *S. solfataricus* P1 proteins can be definitively identified using the P2 genomic database by mapping 420 2-D gel protein spots, representing 224 different gene products. In mapping P1 we identified, two putative enzymes, biotin-acetyl-CoA carboxylase/synthase and cobyrinic acid synthase, whose role in *Sulfolobus* metabolism has yet to be determined. We also compared the functional proteomes of each strain when grown with different carbon sources to see if we could determine the physiological state of *S. solfataricus* that could account for the growth response observed for those carbon sources. When 0.1% tryptone is used as the sole carbon and energy source, both P1 and P2 expressed an abundance of a protein that is currently annotated as an aerobic CO dehydrogenase (CODH) and N-methylhydantoinase, among other proteins. The CODH is most likely an aldehyde oxidoreductase that acts to detoxify the cell from the formaldehyde that is formed from creatine degradation (via N-methylhydantoinase). Further evidence suggests that the aldehyde oxidoreductase may be part of an energy conservation mechanism, as yet unknown. When a mixture of 0.1% tryptone, 0.1% yeast extract, and 0.1% sucrose was used protein biosynthesis was more evident on 2-D gels

relative to using tryptone and growth yields were higher, but the expression of the aldehyde oxidoreductase and N-Methylhydantoinases were significantly down-regulated.

Introduction

Sulfolobales are a genus of ubiquitous archaeal microorganisms and can be found in environments that are both extremely thermophilic and acidophilic (Rice et al, 2001). *Sulfolobus solfataricus* strains P1 (DSM 1616) and P2 (DSM 1617) were originally isolated from two different drainages of the same habitat, the Pisciarelli solfataric hot springs near Naples, Italy (Zillig et al, 1980). P1 was collected from a spring which flowed into a water hole in one drainage, while P2 was isolated from a pool at the end of a nearby second drainage. The temperature of the springs ranged from 74 to 89°C and the pH ranged from 1.4 to 6.0 (De Rosa et al, 1975; Zillig et al, 1980). The two isolates were have similar morphology (including the presence of an S-layer), contain the same DNA GC content, and have very similar high molecular weight components in their DNA-dependent RNA polymerase (using SDS-PAGE) (Zillig et al, 1980). P1 and P2 differ, however, in their DNA restriction fragment patterns, in the mobility of some of the low molecular weight components of their RNA polymerase, in the denatured state of their respective S-layers in SDS and DTT, and in their ability to utilize different carbon sources for growth (Grogan, 1989; Zillig et al, 1980).

We judged a proteomic study of strain P1 to be important because our coworkers at MSU have found it generally more susceptible to certain types of viral infection and

because it was readily apparent from preliminary virus/host experiments that protein expression differences between the two uninfected strains could be observed on 2-D gels. Since only the P2 genome is available, our first goal was to determine if proteins from strain P1 could be identified using the published P2 genomic database. Our second goal was to determine if there were any baseline proteomic differences that could be detected between these two strains by comparing the tris-soluble(aq) and tris-insoluble(aq) protein extracts of P1, using 2-DE, to those of the previously presented P2 gels in Chapter 5. And lastly, we wanted to assess if the striking expression of certain proteins by P1 and P2, such as the annotated carbon monoxide dehydrogenase and N-methylhydantoinase, was specifically determined by the type of carbon source used in the culture media to clarify for future proteomic studies.

Material and Methods

Cell Cultures and Harvesting

Sulfolobus solfataricus P1 (DSM 1616) and P2 (DSM 1617) were obtained from the laboratory of W. Zillig of the Lehrstuhl für Mikrobiologie, Universität Regensburg, Regensburg, Germany. Strains P1 and P2 were both grown aerobically at 80°C and pH 3.2 in modified 9K medium (Schleper et al, 1994) without additional vitamins and P1 was supplemented with a mixture of 0.1% tryptone, 0.1% yeast extract, and 0.1% sucrose (wt/v), while P2 was supplemented with 0.1% tryptone (wt/v), unless otherwise indicated.

2-DE, Spot Detection, Image Analysis, and Database Searches

All materials and methods applied to P1 and P2 were as previously described (Chapters 4 and 5), with the following exceptions: tris-soluble protein extract (300 µg) from P1 was separated only over IPG NL 3-10 and a 8-18% polyacrylamide gel. Database searches for identification of P1 and P2 proteins, using PMF, was performed against the FASTA P2 database.

Results: Proteomic Mapping of P1 Proteins

Sulfolobus solfataricus P1 proteins which were isolated on 2-D gels could be just as easily identified using the P2 database as the P2 proteins. Tris-soluble(aq) and tris-insoluble(aq) protein extracts for P1 were separated using large format (20 X 20 cm) 2-DE and identified with PMF. Tris-soluble(aq) fractions were separated using only IPG NL 3-10 strips and 8-18% polyacrylamide gels, while tris-insoluble(aq) protein extracts were separated over three different IPG strips (IPG 3-10, 4-7, 6-11) and 8-18% gels. The tris-soluble(aq) fraction was initially used to show that one could use the P2 database to identify P1 proteins with confidence. Two hundred and fifty seven spots representing 164 different gene products were identified on the gel obtained from the tris-soluble(aq) fraction (Figure 32) and 163 spots, representing 75 different gene products, were identified on the gels from the tris-insoluble(aq) fraction (Figures 33-35). Overall, a total

of 420 protein spots representing 224 individual gene products were mapped on 2-D gels from P1 using the P2 database.

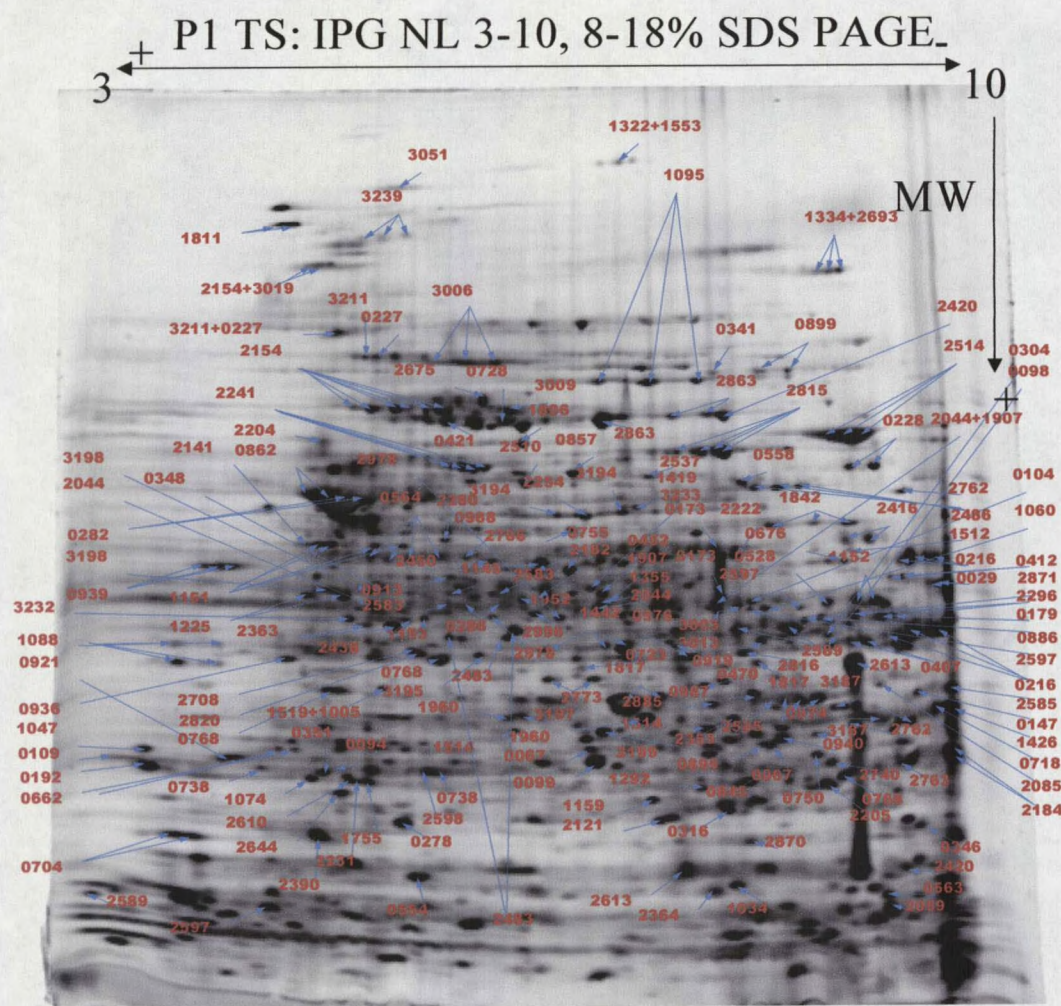


Figure 32. Two-dimensional gel electrophoresis (2-DE) of 300 μ g of tris-soluble(aq) proteins from *Sulfolobus solfataricus* P1. Isoelectric focusing was performed using an IPG NL 3-10 strip. IEF buffer consisted of 8 M urea, 2% CHAPS, 2 mM TBP, in 50 mM Tris-HCl. SDS-PAGE was performed using an 8-18% polyacrylamide gel and Laemmli buffer. Gels were stained with SYPRO Ruby and imaged at 532 nm using the Molecular Imager FX (Bio-Rad). Proteins spots were excised, treated with methyl-modified trypsin, and peptides detected using with MALDI TOF MS. Proteins were identified by PMF using ProteinProbe (Micromass). Spot labeling was performed using Delta2D software (Decodon). Spots are labeled according to a matching *S. solfataricus* P2 number SSOXXXX and are listed in Table 13.

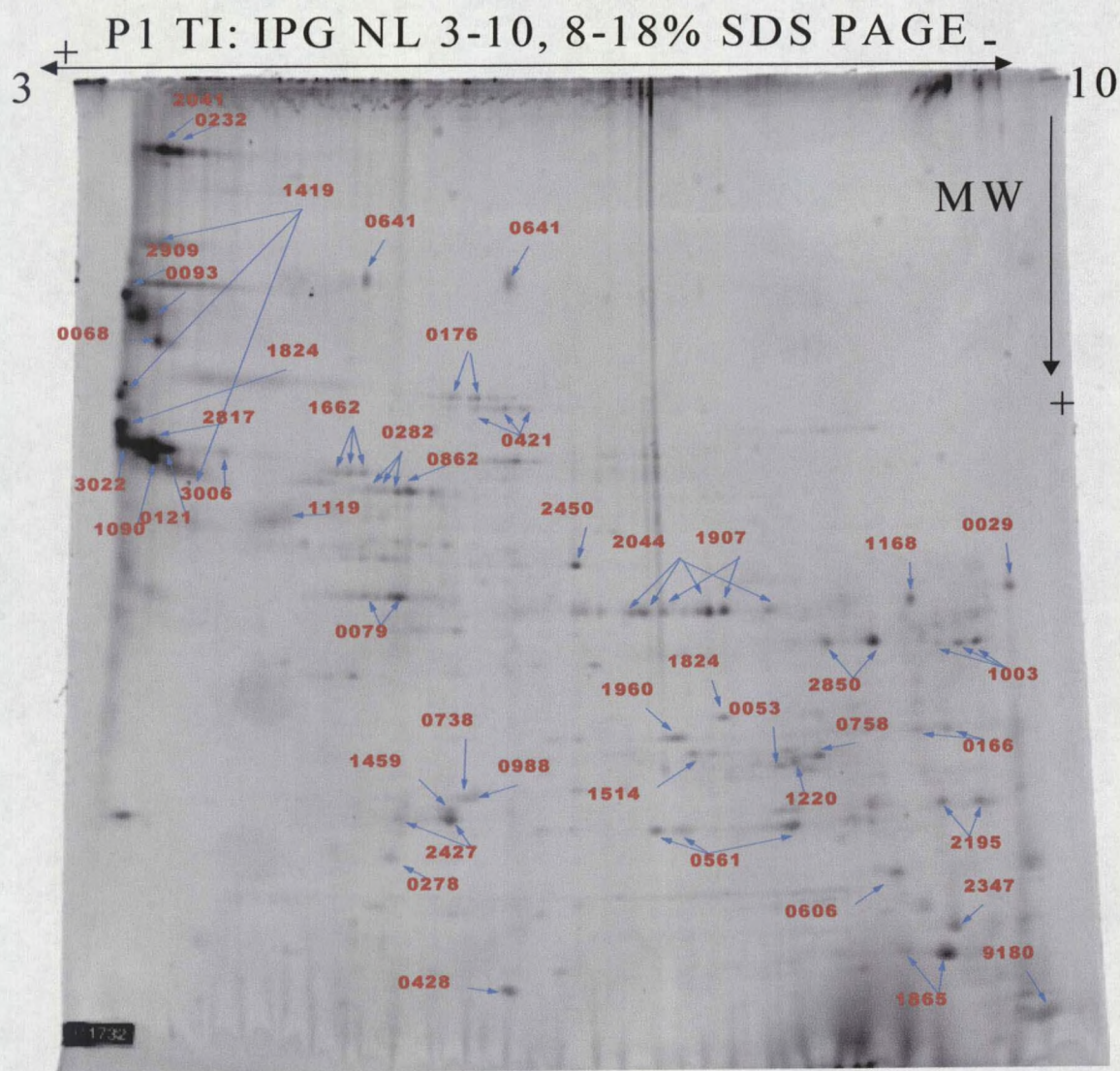


Figure 33. Two-dimensional gel electrophoresis (2-DE) of tris-insoluble(aq) proteins from *Sulfolobus solfataricus* P1. Isoelectric focusing was performed using an IPG NL 3-10 strip. IEF buffer consisted of 5 M urea, 1.4 M thiourea, 3% CHAPS, 1% ASB-14, 2 mM TBP, in 50 mM Tris-HCl. SDS-PAGE was performed using an 8-18% polyacrylamide gel and Laemmli buffer. Gels were stained with SYPRO Ruby and imaged at 532 nm using the Molecular Imager FX (Bio-Rad). Proteins spots were excised, treated with methyl-modified trypsin, and peptides detected using with MALDI TOF MS. Proteins were identified by PMF using Mascot (Matrix Science). Spot labeling was performed using Delta2D software (Decodon). Spots are labeled according to a matching *S. solfataricus* P2 number SSOXXXX and are listed in Table 13.

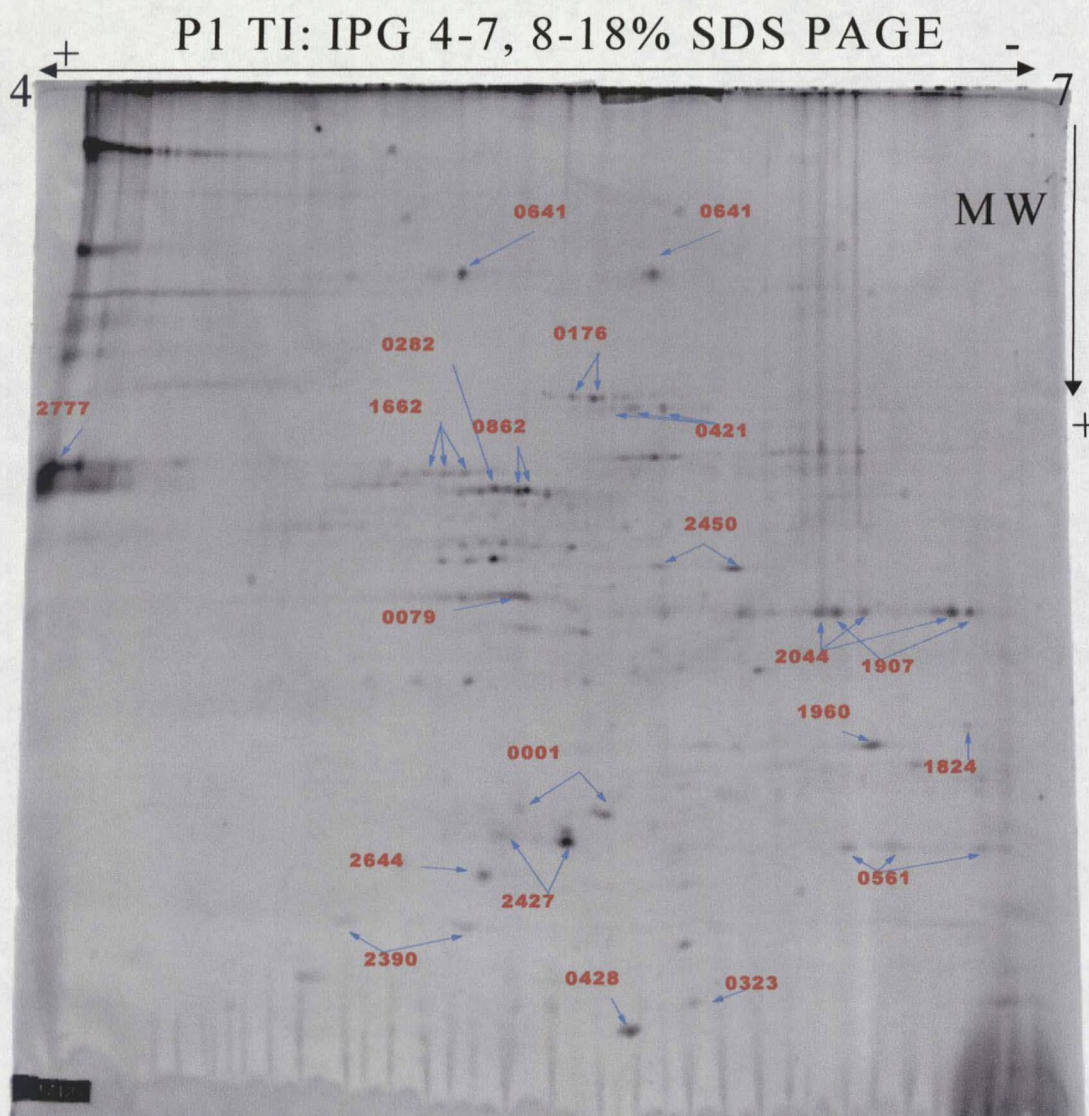


Figure 34. Two-dimensional gel electrophoresis (2-DE) of tris-insoluble(aq) proteins from *Sulfolobus solfataricus* P1. Isoelectric focusing was performed using an IPG NL 4-7 strip. IEF buffer consisted of 5 M urea, 1.4 M thiourea, 3% CHAPS, 1% ASB-14, 2 mM TBP, in 50 mM Tris-HCl. SDS-PAGE was performed using an 8-18% polyacrylamide gel and Laemmli buffer. Gels were stained with SYPRO Ruby and imaged at 532 nm using the Molecular Imager FX (Bio-Rad). Proteins spots were excised, treated with methyl-modified trypsin, and peptides detected using with MALDI TOF MS. Proteins were identified by PMF using Mascot (Matrix Science). Spot labeling was performed using Delta2D software (Decodon). Spots are labeled according to a matching *S. solfataricus* P2 number SSOXXXX and are listed in Table 13.

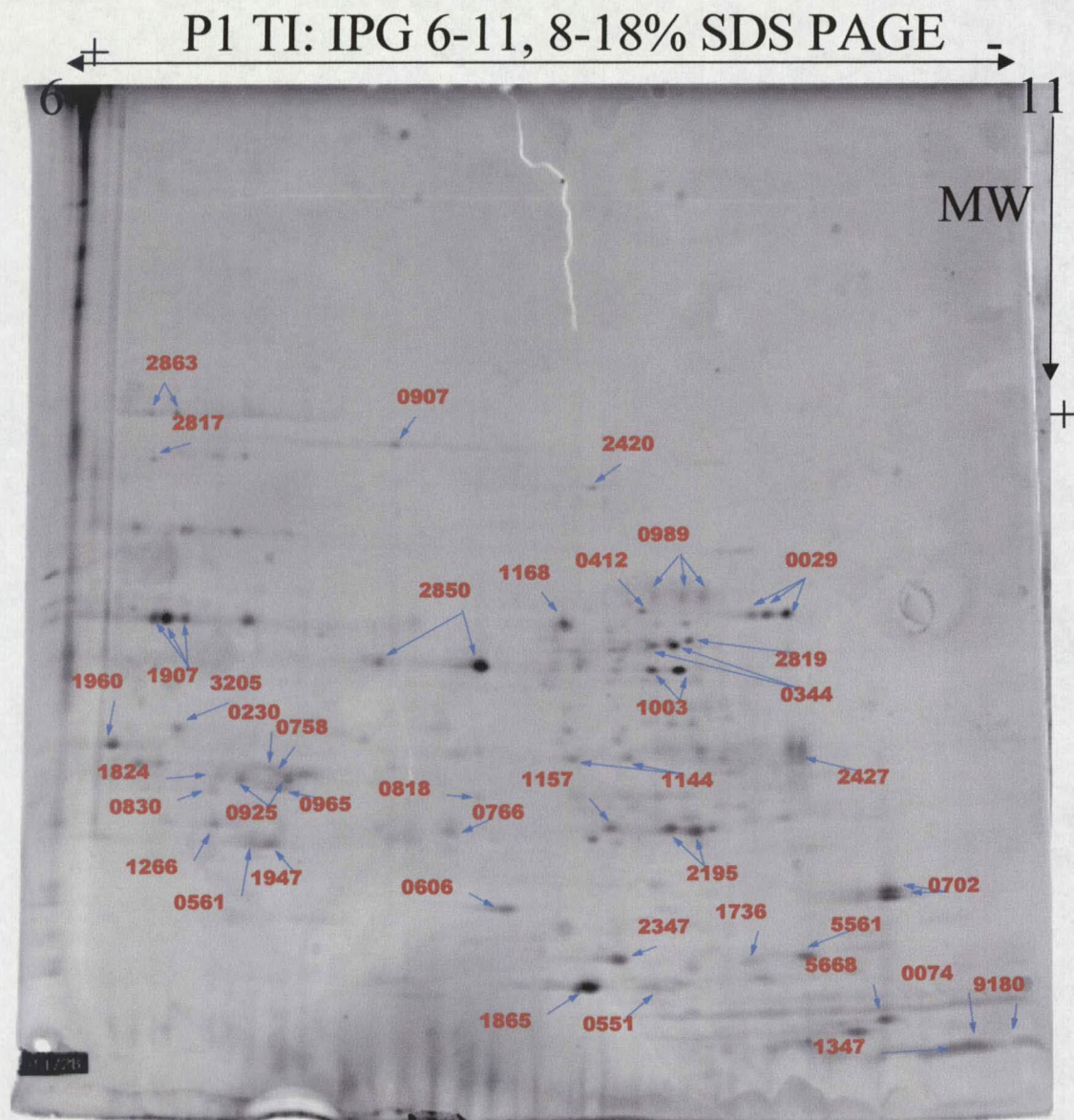


Figure 35. Two-dimensional gel electrophoresis (2-DE) of tris-insoluble(aq) proteins from *Sulfolobus solfataricus* P1. Isoelectric focusing was performed using an IPG NL 6-11 strip. IEF buffer consisted of 5 M urea, 1.4 M thiourea, 3% CHAPS, 1% ASB-14, 2 mM TBP, in 50 mM Tris-HCl. SDS-PAGE was performed using an 8-18% polyacrylamide gel and Laemmli buffer. Gels were stained with SYPRO Ruby and imaged at 532 nm using the Molecular Imager FX (Bio-Rad). Proteins spots were excised, treated with methyl-modified trypsin, and peptides detected using with MALDI TOF MS. Proteins were identified by PMF using Mascot (Matrix Science). Spot labeling was performed using Delta2D software (Decodon). Spots are labeled according to a matching *S. solfataricus* P2 number SSOXXXX and are listed in Table 13.

The proteins identified from strain P1 are listed in Table 13 according to their *Sulfolobus* number (SSOXXXX) along with the matching score and the percent of the protein identified from the match. Those proteins which were identified in the gels obtained from only the tris-soluble fraction are listed with their SSO number in black font, those found only in the tris-insoluble fraction are listed in red, and those which were found in both fraction are labeled in blue (column 1 of Table 13). The score for the protein identified on a particular gel is listed under the IPG range used to obtain the gel and, in a column adjacent to it, the percentage of the protein identified is given. Protein identifications for the tris-soluble IPG 3-10 gels were analyzed using ProteinProbe® (Micromass) and, hence, are scored according to the number of matching peptides found (typed in bold black font, column 2 only of Table 13) and those identified from the tris-insoluble IPG 3-10 gel were matched using a Mascot® MOWSE score (Matrix Science) and are listed in either blue (if the score had $\geq 95\%$ confidence) or red font (if it was $\geq 90\%$ confidence). Proteins identified from 2-D gels obtained using IPG 4-7 and IPG 6-11 strips were also matched using MOWSE scores and are listed in either red or blue font. In a great majority of the cases, the internally calibrated MS spectra allowed for positive protein identifications with a mass accuracy of 100 ppm or less. There were only a handful of cases where good spectra did not produce positive identifications. In those cases, there were no matches which had mass errors less than 200 ppm (most were > 500 ppm) and, therefore, positive identifications could not be made. It is unclear whether post-translational modifications (such as glycosylation) were the cause, or if the peptide masses acquired were distinctly unique for P1 proteins that are not encoded for in the

genome of P2. Additional causes for no matches, which are possible but not probable, include the potential use of alternative start codons by P1 (De Vendittis and Bocchini, 1996; She et al, 2001; Tolstrup et al, 2000) and the presence of extrachromosomal elements which have expressed proteins (Rice et al, 2001; She et al, 2001; Zillig et al, 1998). Similar to the findings of the P2 proteomic mapping in Chapter 5, the P1 proteins identified from the tris-soluble(aq) extracts were mostly cytosolic and metabolism-related proteins and those from aqueous insoluble extract were mostly membrane associated .

Table 13.

P1 Tris-Soluble and Tris-Insoluble Proteins Identified by *S. solfataricus* ORF Number, SSO

SSO# Column				IPG Column						
Black SSO# found only in the tris-soluble fraction				Black = ProteinProbe number of peptides found matching						
Red SSO # found only in tris-insoluble fraction				Red = Mascot Mowse matching score, 90% confidence or greater						
Blue SSO# found in both tris-soluble and tris-insoluble fractions				Blue = Mascot Mowse matching score, 95% confidence or greater						
Pct. Column = percentage of the protein identified with MS										
Type = protein functional category										
AAB: Amino Acid Biosynthesis				IS: Insertion Element						
CEM: Cell Envelope Membrane				Nuc: Nucleotide Synthesis/Metabolism						
CIM: Central Intermediary Metabolism				PPM: Proteases and Protein Modifications						
CF: Cofactor Biosynthesis				TrL: Translation						
CP: Cellular Processes				TrR: Transcription and Regulation						
EM: Energy Metabolism				TrSP: Transport						
LM: Lipid Metabolism				RR: Replication and Repair						
HP Hypothetical Protein				Unc: Uncategorized						
S_sol#	IPG 3-10	Pct	IPG 4-7	Pct	IPG 6-11	Pct	MW	pI	Type	Function
SSO0001	—	—	45	23	—	—	23239	6.67	HP	Hypothetical protein
SSO0029	18	39	—	—	—	—	43433	10.21	HP	Conserved hypothetical protein
SSO0029	133	48	—	—	102	37	43433	10.21	HP	Conserved hypothetical protein
SSO0053	66	39	—	—	—	—	27837	7.20	TrSP	ABC transporter, ATP binding protein
SSO0067	11	49	—	—	—	—	26007	9.04	TRL	(rps2AB) SSU ribosomal protein S2AB
SSO0068	54	47	—	—	—	—	16152	10.91	TRL	(rps9AB) SSU ribosomal protein S9AB
SSO0074	—	—	—	—	51	21	18629	11.09	TRL	(rps13AB) SSU ribosomal protein S13AB
SSO0079	79	40	66	32	—	—	45409	5.00	RR	(dnaG) Bacterial-like DNA primase
SSO0093	60	35	—	—	—	—	66805	8.91	TRL	(glx) Glutamyl-tRNA synthetase
SSO0094	8	49	—	—	—	—	25842	5.30	HP	Conserved hypothetical protein
SSO0098	9	37	—	—	—	—	33252	6.69	PPM	Methionine aminopeptidase 2
SSO0099	11	66	—	—	—	—	25108	6.24	HP	Conserved hypothetical protein
SSO0104	5	22	—	—	—	—	45571	7.82	AAB	(aspB-1) Aspartate aminotransferase
SSO0109	12	48	—	—	—	—	26580	9.58	HP	Conserved hypothetical protein
SSO0121	41	18	—	—	—	—	79965	8.03	HP	Hypothetical protein
SSO0147	13	59	—	—	—	—	34375	9.71	HP	Hypothetical protein
SSO0166	58	43	—	—	—	—	35423	8.35	EM	(galE-2) UDP-glucose 4-epimerase
SSO0173	15	39	—	—	—	—	48769	7.23	TRL	(aspS) Aspartyl-tRNA synthetase
SSO0176	46	15	68	29	—	—	85807	5.75	CP	AAA family ATPase
SSO0179	9	47	—	—	—	—	34332	8.80	RR	(rad2) DNA repair endo/exonuclease
SSO0192	9	38	—	—	—	—	26033	4.70	EM	Glutaredoxin related protein
SSO0216	24	53	—	—	—	—	48489	9.63	TRL	(tuF-1) Elongation factor 1-alpha
SSO0227	11	21	—	—	—	—	73123	6.64	TrR	(rpoB1) DNA-directed RNA polymerase, subunit B'
SSO0228	26	52	—	—	—	—	66933	9.48	TRL	(infB) Translation initiation factor IF2 homolog
SSO0230	—	—	—	—	58	47	15554	8.40	CIM	(ndk) Nucleoside diphosphate kinase (NDP kinase)
SSO0232	30	17	—	—	—	—	52337	7.26	TRL	Glutamyl-tRNA -Gln amidotransferase, subunit B
SSO0278	9	59	—	—	—	—	21376	5.43	PPM	Proteasome subunit
SSO0278	55	32	—	—	—	—	21376	5.43	PPM	Proteasome subunit
SSO0282	21	43	—	—	—	—	60367	5.51	TRL	(thsB) Thermosome beta subunit(thermophilic factor 55)
SSO0282	84	34	51	29	—	—	60367	5.51	TRL	(thsB) Thermosome beta subunit(thermophilic factor 55)
SSO0286	13	53	—	—	—	—	42708	6.12	HP	Conserved hypothetical protein
SSO0304	6	25	—	—	—	—	36265	7.95	AAB	(aroG) Phospho-2-dehydro-3-deoxyheptonate aldolase
SSO0316	5	30	—	—	—	—	24243	7.32	CP	(sod) Superoxide dismutase [Fe]
SSO0323	—	—	53	21	—	—	19549	5.77	EM	(NuoC) NADH dehydrogenase subunit C
SSO0341	10	14	—	—	—	—	102979	7.12	TRL	(alaS) Alanyl-tRNA synthetase
SSO0344	—	—	—	—	88	51	37061	10.02	TRL	(rplp0) LSU acidic ribosomal protein P0 homolog
SSO0346	9	59	—	—	—	—	18524	10.03	TRL	(rpl11AB) LSU ribosomal protein L11AB
SSO0348	8	43	—	—	—	—	29550	7.46	CP	(srpR) Signal recognition particle protein
SSO0351	10	54	—	—	—	—	24373	5.10	TRL	(eIF6) Translation initiation factor 6 (eIF 6)
SSO0407	21	55	—	—	—	—	38250	8.55	AAB	(metE-2) 5-methyltetrahydropteroyltrimethylglutamate
SSO0412	11	29	—	—	—	—	45792	8.48	TRL	(eIF2G) Translation initiation factor eIF 2 gamma

Table 13, continued part 2

P1 Tris-Soluble and Tris-Insoluble Proteins Identified by *S. solfataricus* ORF Number, SSO

SSO# Column							IPG Column			
Black SSO# found only in the tris-soluble fraction							Black = ProteinProbe number of peptides found matching			
Red SSO # found only in tris-insoluble fraction							Red = Mascot Mowse matching score, 90% confidence or greater			
Blue SSO# found in both tris-soluble and tris-insoluble fractions							Blue = Mascot Mowse matching score, 95% confidence or greater			
Pct. Column = percentage of the protein identified with MS										
Type = protein functional category										
AAB: Amino Acid Biosynthesis							IS: Insertion Element			
CEM: Cell Envelope Membrane							Nuc: Nucleotide Synthesis/Metabolism			
CIM: Central Intermediary Metabolism							PPM: Proteases and Protein Modifications			
CF: Cofactor Biosynthesis							TrL: Translation			
CP: Cellular Processes							TrR: Transcription and Regulation			
EM: Energy Metabolism							TrSP: Transport			
LM: Lipid Metabolism							RR: Replication and Repair			
HP Hypothetical Protein							Unc: Uncategorized			
S_sol#	IPG 3-10	Pct	IPG 4-7	Pct	IPG 6-11	Pct	MW	pI	Type	Function
SSO0412	-	-	-	-	47	18	45792	8.48	TRL	(eif2G) Translation initiation factor eIF 2 gamma
SSO0421	37	59	-	-	-	-	86346	6.73	CP	AAA family ATPase
SSO0421	49	33	49	38	-	-	13123	7.04	HP	Hypothetical protein
SSO0428	65	54	47	26	-	-	13123	7.04	HP	Hypothetical protein
SSO0452	16	43	-	-	-	-	45448	6.63	TRL	(trpS) Tryptophanyl-tRNA synthetase
SSO0470	9	46	-	-	-	-	33641	6.71	HP	Conserved hypothetical protein
SSO0528	9	37	-	-	-	-	37595	8.71	EM	(gap) Glyceraldehyde-3-phosphate dehydrogenase
SSO0551	-	-	-	-	53	37	24406	10.10	HP	Conserved hypothetical protein
SSO0554	6	47	-	-	-	-	18154	6.53	TRL	Ribosomal protein L11 methyltransferase
SSO0558	24	43	-	-	-	-	66516	7.44	TRL	(metS) Methionyl-tRNA synthetase
SSO0561	66	47	47	38	82	30	22633	7.32	EM	(atpE) ATP synthase subunit E (EC: 3.6.1.34)
SSO0563	11	28	-	-	-	-	66411	6.30	EM	(atpA) ATP synthase subunit A
SSO0564	22	50	-	-	-	-	51089	5.16	EM	(atpB) ATP synthase subunit B
SSO0576	15	66	-	-	-	-	36811	7.08	AAB	(ilvC-1) Ketol-acid reductoisomerase
SSO0606	60	66	-	-	101	59	17975	9.03	TrR	Transcriptional regulator, putative
SSO0641	51	18	47	11	-	-	118204	5.50	AAB	(carB) Carbamoyl-phosphate synthase large subunit
SSO0662	10	49	-	-	-	-	26757	5.08	CF	(bioR) Biotin-(acetyl-CoA carboxylase) synthetase
SSO0676	19	57	-	-	-	-	43235	9.16	CF	(moaA-1) Molybdopterin biosynthesis protein
SSO0702	-	-	-	-	129	54	20215	10.39	TRL	(rpl6AB) LSU ribosomal protein L6AB
SSO0704	11	56	-	-	-	-	20652	10.94	TRL	(rpl5AB) LSU ribosomal protein L5AB
SSO0718	13	52	-	-	-	-	29534	10.84	TRL	(rpl4AE) LSU ribosomal protein L4AE
SSO0723	10	39	-	-	-	-	37089	6.34	AAB	(leuB) 3-isopropylmalate dehydrogenase
SSO0728	35	60	-	-	-	-	81820	6.49	TRL	(EF-2) Elongation factor 2
SSO0738	8	44	-	-	-	-	26588	5.82	PPM	Proteasome subunit
SSO0738	47	25	-	-	39	20	26588	5.82	PPM	Proteasome subunit
SSO0750	10	40	-	-	-	-	24661	8.86	HP	Conserved hypothetical protein
SSO0755	17	46	-	-	-	-	48937	7.77	AAB	(ahcY) S-adenosyl-L-homocysteine hydrolase
SSO0758	51	22	-	-	-	-	26716	7.52	PPM	Peptidyl-prolyl cis-trans isomerase
SSO0766	9	44	-	-	-	-	24903	9.45	PPM	Proteasome subunit
SSO0766	-	-	-	-	47	12	24903	9.45	PPM	Proteasome subunit
SSO0768	19	68	-	-	-	-	37740	5.68	RR	(rfc) Activator 1, replication factor C, small subunit
SSO0818	-	-	-	-	49	29	28126	9.12	HP	Conserved hypothetical protein
SSO0830	-	-	-	-	30	24	36486	7.05	EM	(ribB-1) dTDP-Glucose 4,6-dehydratase
SSO0845	16	63	-	-	-	-	27012	7.10	HP	Hypothetical protein
SSO0857	12	28	-	-	-	-	71386	6.85	TRL	(argS) Arginyl-tRNA synthetase
SSO0862	18	48	-	-	-	-	59674	5.22	TRL	(thsA) Thermosome alpha subunit (thermophilic factor 55)
SSO0862	45	16	48	18	-	-	59674	5.22	TRL	(thsA) Thermosome alpha subunit (thermophilic factor 55)
SSO0886	11	43	-	-	-	-	39377	8.08	HP	Conserved hypothetical protein
SSO0895	8	44	-	-	-	-	28588	7.14	AAB	(trpC) Indole-3-glycerol phosphate synthase
SSO0899	15	19	-	-	-	-	98577	8.37	TRL	(valS) Valyl-tRNA synthetase
SSO0907	-	-	-	-	51	31	76573	10.32	RR	(topA) DNA topoisomerase I family A

Table 13, continued part 3

P1 Tris-Soluble and Tris-Insoluble Proteins Identified by *S. solfataricus* ORF Number, SSO

SSO# Column				IPG Column						
Black SSO# found only in the tris-soluble fraction				Black = ProteinProbe number of peptides found matching						
Red SSO # found only in tris-insoluble fraction				Red = Mascot Mowse matching score, 90% confidence or greater						
Blue SSO# found in both tris-soluble and tris-insoluble fractions				Blue = Mascot Mowse matching score, 95% confidence or greater						
Pct. Column = percentage of the protein identified with MS										
Type = protein functional category										
AAB: Amino Acid Biosynthesis				IS: Insertion Element						
CEM: Cell Envelope Membrane				Nuc: Nucleotide Synthesis/Metabolism						
CIM: Central Intermediary Metabolism				PPM: Proteases and Protein Modifications						
CF: Cofactor Biosynthesis				TrL: Translation						
CP: Cellular Processes				TrR: Transcription and Regulation						
EM: Energy Metabolism				TrSP: Transport						
LM: Lipid Metabolism				RR: Replication and Repair						
HP Hypothetical Protein				Unc: Uncategorized						
S_sol#	IPG 3-10	Pct	IPG 4-7	Pct	IPG 6-11	Pct	MW	pI	Type	Function
SSO0913	11	43	--	--	--	--	46745	5.38	EM	Enolase
SSO0919	13	49	--	--	--	--	39724	6.97	EM	Aminomethyltransferase
SSO0921	6	34	--	--	--	--	28178	6.70	HP	Conserved hypothetical protein
SSO0925	--	--	--	--	68	38	17769	9.01	TrSP	ABC transporter
SSO0936	4	8	--	--	--	--	71533	5.91	TRL	Glutamyl-tRNA -Gln amidotransferase, subunit B
SSO0939	18	59	--	--	--	--	46692	10.32	TrR	Pre mRNA splicing protein
SSO0940	10	57	--	--	--	--	26422	8.08	TrR	Fibrillar-in-like pre-rRNA processing protein
SSO0965	--	--	--	--	62	30	108518	10.02	Unc	ATP-dependent helicase
SSO0967	17	69	--	--	--	--	35050	7.87	Unc	Deoxyhypusine synthase
SSO0974	15	64	--	--	--	--	31658	9.59	Unc	3-hydroxyisobutyrate dehydrogenase
SSO0988	17	37	--	--	--	--	53641	5.75	EM	Alpha amylase
SSO0988	51	28	--	--	--	--	53641	5.75	EM	Alpha amylase
SSO0989	--	--	--	--	53	22	48225	9.86	CEM	Sugar phosphate nucleotydyl transferase
SSO1003	48	22	--	--	78	33	36800	9.54	TrSP	Sugar ABC transporter
SSO1005	6	21	--	--	--	--	32909	5.37	HP	Conserved hypothetical protein
SSO1034	4	20	--	--	--	--	23556	8.35	HP	ABC transporter, ATP binding protein.
SSO1047	12	58	--	--	--	--	27567	4.36	RR	(pcnA-2) Proliferating cell nuclear antigen homolog
SSO1060	7	30	--	--	--	--	31829	7.11	CF	(lplA-like1) Lipoate protein ligase A, hypothetical
SSO1074	10	60	--	--	--	--	26089	5.10	HP	Conserved hypothetical protein
SSO1088	7	45	--	--	--	--	23232	4.74	HP	Conserved hypothetical protein
SSO1090	35	24	--	--	--	--	30554	4.64	CP	Serine/threonine phosphatase type 1
SSO1095	25	44	--	--	--	--	95774	6.58	EM	Aconitate hydratase
SSO1119	47	21	--	--	--	--	46460	4.72	HP	Hypothetical protein
SSO1144	--	--	--	--	66	45	31650	9.32	TrSP	ABC transporter, ATP binding protein
SSO1145	12	33	--	--	--	--	47778	8.72	AAB	(trpB-like) Tryptophan synthase beta chain, hypothetical
SSO1151	27	71	--	--	--	--	51710	4.99	HP	Conserved hypothetical protein
SSO1152	16	51	--	--	--	--	46122	9.50	HP	Conserved hypothetical protein
SSO1153	11	33	--	--	--	--	41473	6.84	LM	(acd-1) Acyl-CoA dehydrogenase
SSO1157	--	--	--	--	41	12	23712	9.98	HP	Conserved hypothetical protein
SSO1159	8	48	--	--	--	--	23009	7.52	Unc	Haloacetate dehalogenase
SSO1168	51	24	--	--	86	57	41884	9.26	TrSP	Sugar ABC transporter
SSO1220	36	24	--	--	--	--	39682	7.30	EM	(adh-3) Alcohol dehydrogenase (Zn containing)
SSO1225	15	46	--	--	--	--	45525	8.32	CP	(todA) Toluene 1,2-dioxygenase system ferredoxin--NAD(+) re
SSO1266	--	--	--	--	32	25	32335	6.35	TrSP	ABC transporter, ATP binding protein
SSO1292	11	59	--	--	--	--	21295	7.35	HP	Hypothetical protein
SSO1322	10	37	--	--	--	--	37185	5.44	AAB	(ilvC-2) Ketol-acid reductoisomerase
SSO1334	12	15	--	--	--	--	94701	7.25	EM	(aceB/mas) Malate synthase, putative
SSO1347	--	--	--	--	55	36	21819	11.15	ISE	Third ORF in transposon ISC1904
SSO1355	11	25	--	--	--	--	43068	6.31	PPM	(cpsA-1) Thermostable carboxypeptidase
SSO1419	22	42	--	--	--	--	61784	5.45	PPM	(apeH-1) Acylaminoacyl-peptidase, putative
SSO1419	38	24	--	--	--	--	61784	5.45	PPM	(apeH-1) Acylaminoacyl-peptidase, putative

Table 13, continued part 4

P1 Tris-Soluble and Tris-Insoluble Proteins Identified by *S. solfataricus* ORF Number, SSO

SSO# Column							IPG Column			
Black SSO# found only in the tris-soluble fraction							Black = ProteinProbe number of peptides found matching			
Red SSO # found only in tris-insoluble fraction							Red = Mascot Mowse matching score, 90% confidence or greater			
Blue SSO# found in both tris-soluble and tris-insoluble fractions							Blue = Mascot Mowse matching score, 95% confidence or greater			
Pct. Column = percentage of the protein identified with MS										
Type = protein functional category										
AAB: Amino Acid Biosynthesis							IS: Insertion Element			
CEM: Cell Envelope Membrane							Nuc: Nucleotide Synthesis/Metabolism			
CIM: Central Intermediary Metabolism							PPM: Proteases and Protein Modifications			
CF: Cofactor Biosynthesis							TrL: Translation			
CP: Cellular Processes							TrR: Transcription and Regulation			
EM: Energy Metabolism							TrSP: Transport			
LM: Lipid Metabolism							RR: Replication and Repair			
HP Hypothetical Protein							Unc: Uncategorized			
S.sol#	IPG 3-10	Pct	IPG 4-7	Pct	IPG 6-11	Pct	MW	pI	Type	Function
SSO1442	12	52	--	--	--	--	35264	6.57	HP	Conserved hypothetical protein
SSO1459	9	27	--	--	--	--	65715	9.49	RR	(dpo2) DNA polymerase II (DNA polymerase B2) C-end
SSO1459	58	32	--	--	--	--	65715	9.49	RR	(dpo2) DNA polymerase II (DNA polymerase B2) carboxy-end
SSO1512	16	36	--	--	--	--	54120	8.00	HP	Hypothetical protein
SSO1514	12	71	--	--	--	--	29872	7.35	HP	Conserved hypothetical protein
SSO1514	47	18	--	--	--	--	29872	7.35	HP	Conserved hypothetical protein
SSO1519	6	21	--	--	--	--	31374	5.22	Nuc	(deoD) Purine nucleoside phosphorylase
SSO1532	10	38	--	--	--	--	39215	6.82	HP	Conserved hypothetical protein
SSO1662	39	27	32	22	--	--	63396	5.28	AAB	(huyB-1) N-methylhydantoinase B
SSO1736	--	--	--	--	32	27	12526	9.39	HP	Conserved hypothetical protein
SSO1755	7	40	--	--	--	--	25160	5.69	HP	Conserved hypothetical protein
SSO1806	27	48	--	--	--	--	72695	6.21	HP	Conserved hypothetical protein
SSO1811	18	38	--	--	--	--	44087	5.77	PPM	(pepQ-like1) Prolidase (Xaa-Pro dipeptidase)
SSO1817	19	52	--	--	--	--	33655	6.90	CIM	(cysA-2) Thiosulfate sulfurtransferase
SSO1824	63	35	51	30	30	11	49215	6.69	HP	Conserved hypothetical protein
SSO1842	17	55	--	--	--	--	51603	8.92	EM	(gapN-2) Glyceraldehyde-3-phosphate dehydrogenase, NADP d
SSO1865	51	51	--	--	45	30	15226	9.02	HP	Conserved hypothetical protein
SSO1907	11	37	--	--	--	--	46031	6.98	AAB	(gdhA-2) NAD specific glutamate dehydrogenase
SSO1907	91	33	60	48	57	48	46031	6.98	AAB	(gdhA-2) NAD specific glutamate dehydrogenase
SSO1947	--	--	--	--	34	27	17105	6.57	HP	Hypothetical protein
SSO1952	8	25	--	--	--	--	42992	6.31	PPM	(cpsA-2) Thermostable carboxypeptidase
SSO1960	18	75	--	--	--	--	31706	6.82	HP	Hypothetical protein
SSO1960	58	48	47	30	123	59	31706	6.82	HP	Hypothetical protein
SSO1991	17	20	--	--	--	--	121639	7.83	HP	Hypothetical protein
SSO2041	25	12	--	--	--	--	57091	6.55	LM	(acsA-5) Acetyl-CoA synthetase (acetate-CoA ligase)
SSO2044	16	47	--	--	--	--	46063	6.98	AAB	(gdhA-4) NAD specific glutamate dehydrogenase
SSO2044	71	39	48	18	--	--	46063	6.98	AAB	(gdhA-4) NAD specific glutamate dehydrogenase
SSO2059	9	25	--	--	--	--	70115	8.10	LM	(acsA-6) Acetyl-CoA synthetase (acetate-CoA ligase)
SSO2085	11	41	--	--	--	--	35688	8.93	HP	Hypothetical protein
SSO2121	13	52	--	--	--	--	24745	7.55	CP	(bcp-2) Peroxiredoxin, bacterioferritin comigratory protein homolog
SSO2141	29	60	--	--	--	--	64282	6.30	PPM	(apeH-2) Acylaminoacyl-peptidase, putative
SSO2154	24	36	--	--	--	--	90515	5.64	PPM	Tricorn protease interacting factor F2
SSO2182	12	36	--	--	--	--	46450	6.73	EM	(idh) Isocitrate dehydrogenase, probable
SSO2184	5	20	--	--	--	--	44554	8.55	RR	(cdc6-3) Cell division control 6/orc1 protein homolog
SSO2195	41	35	--	--	71	37	29663	9.91	CEM	Erythrocyte band 7 membrane protein homolog
SSO2199	20	32	--	--	--	--	118727	5.35	HP	Conserved hypothetical protein
SSO2204	18	30	--	--	--	--	75017	5.75	HP	Conserved hypothetical protein
SSO2205	9	51	--	--	--	--	25945	6.53	LM	(fabG-4) 3-oxoacyl-(acyl carrier protein) reductase
SSO2222	15	61	--	--	--	--	36542	9.28	Nuc	(trxB-1) Thioredoxin reductase
SSO2231	9	51	--	--	--	--	24968	5.47	HP	Hypothetical protein
SSO2241	26	48	--	--	--	--	68436	5.41	Unc	(bps2) BPS2 protein homolog
SSO2254	18	55	--	--	--	--	54308	6.55	HP	Conserved hypothetical protein

Table 13, continued part 5

P1 Tris-Soluble and Tris-Insoluble Proteins Identified by *S. solfataricus* ORF Number, SSO

SSO# Column				IPG Column						
Black SSO# found only in the tris-soluble fraction				Black = ProteinProbe number of peptides found matching						
Red SSO # found only in tris-insoluble fraction				Red = Mascot Mowse matching score, 90% confidence or greater						
Blue SSO# found in both tris-soluble and tris-insoluble fractions				Blue = Mascot Mowse matching score, 95% confidence or greater						
Pct. Column = percentage of the protein identified with MS										
Type = protein functional category										
AAB: Amino Acid Biosynthesis				IS: Insertion Element						
CEM: Cell Envelope Membrane				Nuc: Nucleotide Synthesis/Metabolism						
CIM: Central Intermediary Metabolism				PPM: Proteases and Protein Modifications						
CF: Cofactor Biosynthesis				TrL: Translation						
CP: Cellular Processes				TrR: Transcription and Regulation						
EM: Energy Metabolism				TrSP: Transport						
LM: Lipid Metabolism				RR: Replication and Repair						
HP Hypothetical Protein				Unc: Uncategorized						
S.sol#	IPG 3-10	Pct	IPG 4-7	Pct	IPG 6-11	Pct	MW	pI	Type	Function
SSO2280	19	44	-	-	-	-	55326	6.77	TRL	(cysS) CysteinyI-tRNA synthetase
SSO2296	4	23	-	-	-	-	25548	8.38	CF	(cblE) Cobalamin biosynthesis precomin-6B methylase, putative
SSO2347	41	36	-	-	37	44	17786	9.46	TrR	(asnC) Transcriptional regulatory protein, AsnC family
SSO2353	10	40	-	-	-	-	44951	7.50	LM	Geranylgeranyl hydrogenase, putative
SSO2363	11	33	-	-	-	-	43683	6.18	HP	Conserved hypothetical protein
SSO2364	5	45	-	-	-	-	16138	9.82	RR	(SSB) Single-stranded DNA binding protein
SSO2390	7	40	-	-	-	-	19694	4.78	EM	(ppa) Inorganic pyrophosphatase, putative
SSO2390	-	-	52	32	-	-	19694	4.78	EM	(ppa) Inorganic pyrophosphatase, putative
SSO2416	13	47	-	-	-	-	35221	9.05	Nuc	(trxB-3) Thioredoxin reductase
SSO2420	17	31	-	-	-	-	68396	9.19	CP	AAA family ATPase
SSO2420	-	-	-	-	37	21	68396	9.19	CP	AAA family ATPase
SSO2427	90	60	61	43	56	36	20097	5.10	CP	Small heat shock protein hsp20 family
SSO2438	6	25	-	-	-	-	43139	5.18	HP	Conserved hypothetical protein
SSO2450	14	34	-	-	-	-	53201	8.11	TrR	TATA binding protein (TBP)-interacting protein
SSO2450	47	21	55	26	-	-	53201	8.11	TrR	TATA binding protein (TBP)-interacting protein
SSO2483	10	36	-	-	-	-	37354	5.50	EM	(sucC) Succinyl-CoA synthetase, beta subunit
SSO2486	15	35	-	-	-	-	63230	7.03	TRL	(thrS) Threonyl-tRNA synthetase
SSO2510	21	39	-	-	-	-	62390	6.93	LM	(alkK-2) Medium-chain-fatty-acid-CoA ligase
SSO2514	23	46	-	-	-	-	73038	8.29	LM	3-hydroxyacyl-CoA dehydrogenase/enoyl CoA hydratase
SSO2537	16	35	-	-	-	-	69056	6.77	EM	Phosphoenolpyruvate carboxykinase (GTP)
SSO2583	18	49	-	-	-	-	44686	6.76	LM	(sqdB) Sulfolipid biosynthesis protein
SSO2585	15	64	-	-	-	-	36394	8.86	EM	(sqdB) L-lactate dehydrogenase
SSO2589	16	45	-	-	-	-	42752	7.86	EM	Citrate synthase
SSO2597	14	58	-	-	-	-	41976	7.91	AAB	(agxT) Serine-pyruvate aminotransferase
SSO2598	7	31	-	-	-	-	26668	5.54	TrR	(tenA-2) Transcriptional activator
SSO2610	5	36	-	-	-	-	25586	5.29	HP	Hypothetical protein
SSO2613	13	72	-	-	-	-	17461	8.05	CP	(bcp-4) Peroxiredoxin, bacterioferitin comigratory protein homolog
SSO2644	9	49	-	-	-	-	22420	5.53	HP	Hypothetical protein
SSO2644	-	-	55	36	-	-	22420	5.53	HP	Hypothetical protein
SSO2675	21	32	-	-	-	-	90142	5.56	PPM	Tricorn protease interacting factor F3
SSO2693	16	25	-	-	-	-	67147	6.30	PPM	(apeH-3) Acylaminoacyl-peptidase, putative
SSO2708	8	42	-	-	-	-	38182	5.28	PPM	Hydrogenase expression/formation factor related protein
SSO2740	8	50	-	-	-	-	27595	8.92	HP	Conserved hypothetical protein
SSO2762	11	59	-	-	-	-	30620	10.15	EM	(etfA) Electron transfer flavoprotein, subunit alpha
SSO2763	8	40	-	-	-	-	26560	7.08	EM	(etfB) Electron transfer flavoprotein, subunit beta
SSO2766	21	55	-	-	-	-	52041	6.60	HP	Conserved hypothetical protein
SSO2773	17	71	-	-	-	-	34592	5.94	CIM	(hpcE-2) 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase
SSO2777	-	-	48	25	-	-	13656	5.10	HP	Hypothetical protein
SSO2815	26	53	-	-	-	-	70228	6.82	EM	2-oxoacid-ferredoxin oxidoreductase, alpha chain
SSO2816	8	43	-	-	-	-	35030	8.20	EM	2-oxoacid-ferredoxin oxidoreductase, beta chain
SSO2817	38	12	-	-	47	25	67141	6.31	EM	(etfAB/fixAB) Electron transfer flavoprotein alpha and beta-subunit

Table 13, continued part 6

P1 Tris-Soluble and Tris-Insoluble Proteins Identified by *S. solfataricus* ORF Number, SSO

SSO# Column						IPG Column				
Black SSO# found only in the tris-soluble fraction						Black = ProteinProbe number of peptides found matching				
Red SSO# found only in tris-insoluble fraction						Red = Mascot Mowse matching score, 90% confidence or greater				
Blue SSO# found in both tris-soluble and tris-insoluble fractions						Blue = Mascot Mowse matching score, 95% confidence or greater				
Pct. Column = percentage of the protein identified with MS										
Type = protein functional category										
AAB: Amino Acid Biosynthesis						IS: Insertion Element				
CEM: Cell Envelope Membrane						Nuc: Nucleotide Synthesis/Metabolism				
CIM: Central Intermediary Metabolism						PPM: Proteases and Protein Modifications				
CF: Cofactor Biosynthesis						TrL: Translation				
CP: Cellular Processes						TrR: Transcription and Regulation				
EM: Energy Metabolism						TrSP: Transport				
LM: Lipid Metabolism						RR: Replication and Repair				
HP Hypothetical Protein						Unc: Uncategorized				
S_sol#	IPG 3-10	Pct	IPG 4-7	Pct	IPG 6-11	Pct	MW	pI	Type	Function
SSO2819	-	-	-	-	58	35	43305	9.64	EM	(fixC) FIXC protein homolog
SSO2820	12	46	-	-	-	-	34975	5.44	EM	(ppsA-2) Phosphoenolpyruvate synthase
SSO2850	60	44	-	-	106	37	39113	8.85	TrSP	ABC transporter, ATP binding protein (glucose)
SSO2863	20	35	-	-	-	-	75799	6.89	LM	(acsA-9) Acetyl-CoA synthetase
SSO2863	-	-	-	-	34	14	75799	6.89	LM	(acsA-9) Acetyl-CoA synthetase
SSO2870	8	59	-	-	-	-	18139	6.11	HP	Conserved hypothetical protein
SSO2871	7	19	-	-	-	-	41478	7.28	LM	(hdb-2) 3-hydroxyacyl-CoA dehydrogenase
SSO2885	8	43	-	-	-	-	35080	6.82	EM	Oxidoreductase
SSO2909	54	13	-	-	-	-	72828	6.79	CIM	(cysI) Sulfite reductase hemoprotein beta component
SSO2978	16	63	-	-	-	-	40339	6.83	TRL	(eIF2B) Translation initiation factor eIF-2B, subunit 1
SSO2996	13	42	-	-	-	-	43166	5.93	LM	(hdb-3) 3-hydroxyacyl-CoA dehydrogenase
SSO3003	14	54	-	-	-	-	40891	7.10	EM	(dhg-1) Glucose 1-dehydrogenase
SSO3006	30	38	-	-	-	-	111159	6.12	CIM	Alpha-mannosidase
SSO3006	32	9	-	-	-	-	111159	6.12	CIM	Alpha-mannosidase
SSO3009	15	21	-	-	-	-	79542	6.41	EM	Carbon monoxide dehydrogenase, large chain
SSO3013	13	54	-	-	-	-	36095	7.50	EM	Dehydrogenase, putative
SSO3019	5	12	-	-	-	-	56691	6.47	CIM	(lacS) Beta-glycosidase
SSO3022	27	15	-	-	-	-	84485	5.25	CIM	(xylS) Alpha-xylosidase
SSO3051	32	50	-	-	-	-	80441	5.61	CIM	(malA) Alpha-glucosidase
SSO3187	11	47	-	-	-	-	33812	8.10	AAB	(serA-2) D-3-phosphoglycerate dehydrogenase
SSO3194	20	48	-	-	-	-	56926	6.16	EM	(gapN-3) Glyceraldehyde-3-phos dehydrogenase, NADP dep.
SSO3195	9	42	-	-	-	-	34874	6.10	EM	Fructokinase
SSO3197	6	36	-	-	-	-	34795	7.03	EM	(eda) 2-keto-3-deoxy gluconate aldolase
SSO3198	14	49	-	-	-	-	44729	5.36	CIM	Muconate cycloisomerase related protein
SSO3205	-	-	-	-	47	25	31257	7.44	HP	Conserved hypothetical protein
SSO3211	11	32	-	-	-	-	46521	6.16	AAB	(gabT-2) 4-aminobutyrate aminotransferase
SSO3232	8	35	-	-	-	-	41382	5.34	Nuc	(guaA) GMP synthase, PP-ATPase domain
SSO3233	10	26	-	-	-	-	52327	8.32	CF	Cobyrinic acid synthase
SSO3239	8	16	-	-	-	-	79023	6.12	EM	(cutA-8) Carbon monoxide dehydrogenase, large chain
SSO5561	-	-	-	-	40	37	9628	9.88	TRL	(rps17E) SSU ribosomal protein S17E
SSO5668	-	-	-	-	71	61	10606	10.41	TRL	(rplX) LSU ribosomal protein LX
SSO9180	42	46	-	-	36	30	7279	10.59	Unc	(ssh7A /Sso7d-2) 7 KD DNA-binding protein

Results: Functional Proteome Comparisons Between P1 and P2

In order to determine if any baseline proteomic differences could be detected between P1 and P2, 2-D gel comparisons were performed between the newly mapped P1 gels and the previously mapped P2 gels of Chapter 5. Initial experiments were performed in which P1 and P2 were both cultured using 0.1% tryptone as the sole carbon and energy source, but P1 could only be grown to very low cell densities before the pH of the media increased to the point of P1 cell lysis (OD_{650} of 0.22 for P1 compared to 0.60 for P2) (data not shown). Since P1 could be grown to similar cell densities as P2 (OD_{650} of 0.60) when a mixture of 0.1% tryptone, 0.1% yeast extract, and 0.1% sucrose was used *and* because P1 was found to be more susceptible to viral infection than P2 when using this mixture of carbon sources, we decided to investigate the proteomic differences that may exist between P1 and P2 when each of them were cultured to the same OD. We were primarily interested in identifying proteins which were predominantly expressed by one strain and not the other as influenced by the carbon source. The predominantly expressed proteins of one strain as compared to the other strain are referred to as “unique” in the remainder of the text. Image analysis software (PDQuest 7.0, Bio-Rad) was used to normalize the spot quantities on each gel relative to the total spot quantity on that gel, followed by difference gel analysis to reveal which 2-DE spots were uniquely distinct to P1 gels or P2 gels. The identified spots were excised, subjected to proteolysis by trypsin, and the proteins identified using PMF.

P1 was cultured using a mixture of 0.1% tryptone, 0.1% yeast extract, and 0.1% sucrose, while P2 was grown with only 0.1% tryptone as the carbon and energy source. We were able to detect 28 unique protein spots on the P1 gels and identify 22 of them. Most of the spots (26/28) were detected on the gels obtained from the tris-soluble fraction(aq) (TS gels) and two were detected from gels obtained from the tris-insoluble(aq) fraction (TI gels). From the twenty spots identified from the P1 TS gels, nine were detected only on the P1 gels, another nine were determined have greater than 5-fold abundance relative to their counterpart on the P2 gels, and two were attributed to pI shifts on the P1 gels as compared to the P2 gels (SSO0109, a hypothetical protein and SSO2486, an isoform of threonyl-tRNA synthetase) (Figures 36 and 37). The two spots identified on the P1 TI gels were phosphorylated and nonphosphorylated carbamoyl-phosphate synthase (SSO0641) (Figure 36). A list of the P1 proteins identified along with their PMF scores is given in Table 14.

P2 was cultured with 0.1% tryptone as the sole carbon and energy source and the P2 TS gels contained 14 unique spots, including ten detected only on the P2 gels, two which were significantly up-regulated relative to the P1 gels, and two which could not be identified. No unique spots were detected on P2 TI gels (Figure 36 and 37). The unique proteins detected from P2 TS gels included multiple isoforms of carbon monoxide dehydrogenase (SSO2639), subunits A and B of N-methylhydantoinase, a homologous FIXC electron transfer flavoprotein, a unique AAA ATPase, an isochorismatase related protein, a carboxypeptidase, and three hypothetical proteins, one of which is homologous

to a GTP binding protein (Table 14). The biochemical pathways for each of the P2 proteins noted is given in Table 15

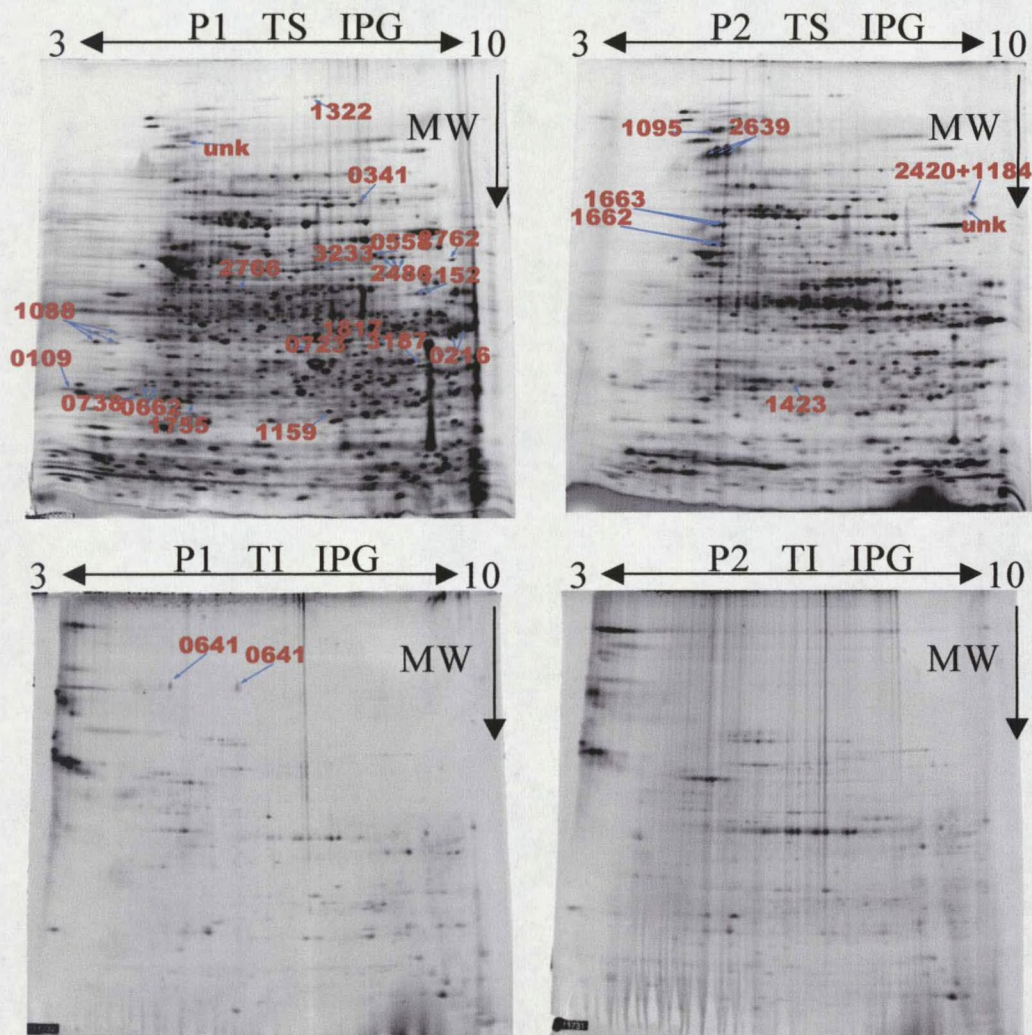


Figure 36. Comparison of 300 μ g of tris-soluble(aq) P1 with 200 μ g of tris-soluble(aq) P2 (TS, top) and equal amounts of tris-insoluble P1 and P2 (TI, bottom) separations over IPG NL 3-10 and 8-18% SDS-PAGE of P1 with P2. P1 was cultured using 0.1% tryptone, 0.1% yeast extract, and 0.1% sucrose, while P2 was supplemented with 0.1% tryptone as the sole carbon and energy source. Gels were stained with SYPRO Ruby and imaged at 532 nm using the Molecular Imager FX (Bio-Rad). Gel images were normalized relative to their total spot quantity and those spots uniquely detected on each gel were identified by PMF and labeled according to their matching *S. solfataricus* P2 number SSOXXXX. A complete list of the uniquely detected proteins from P1 and P2 can be found in Table 14.

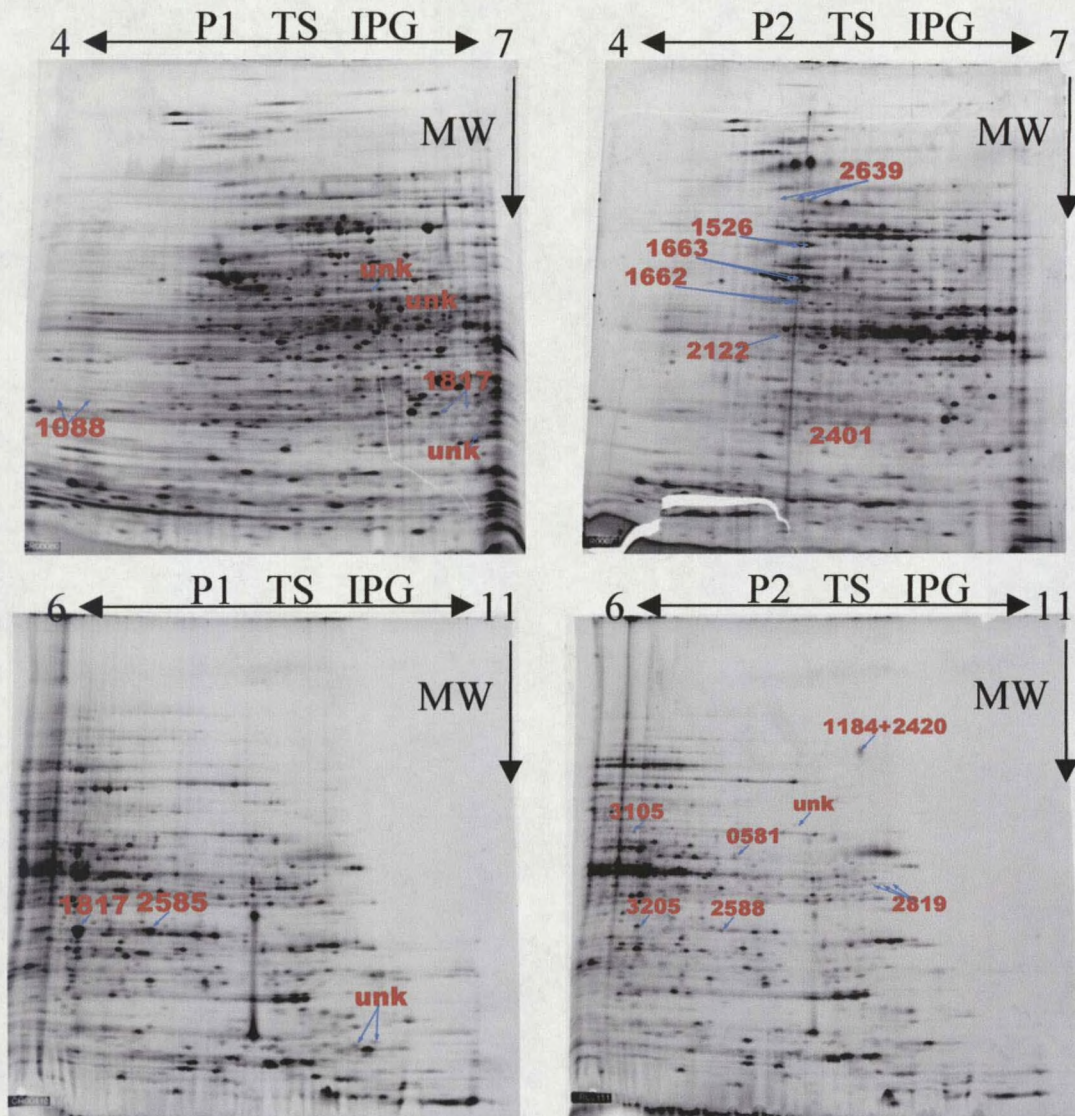


Figure 37. Comparison of 300 μ g of tris-soluble(aq) P1 with 200 μ g of tris-soluble(aq) P2 over IPG 4-7/8-18% SDS-PAGE (TS, top) and IPG 6-11/8-18% SDS-PAGE (TS, bottom). P1 was cultured using 0.1% tryptone, 0.1% yeast extract, and 0.1% sucrose, while P2 was supplemented with 0.1% tryptone as the sole carbon and energy source. Gels were stained with SYPRO Ruby and imaged at 532 nm using the Molecular Imager FX (Bio-Rad). Gel images were normalized relative to their total spot quantity and those spots uniquely detected on each gel were identified by PMF and labeled according to their matching *S. solfataricus* P2 number SSOXXXX. A complete list of the uniquely detected proteins from P1 and P2 can be found in Table 14.

Table 14.

P2 Difference Gel Image Analysis Results, Proteins Uniquely Detected in P1 and P2

SSO# Column				IPG Column						
Black SSO# was found only in the strain identified				Black = ProteinProbe number of peptides found matching						
Red SSO # is expressed in both strains but is highly up-regulated in the strain identified				Blue = Mascot Mowse matching score, 95% confidence or greater						
				Pct = percentage of the protein identified with MS						
Pct. Column = percentage of the protein identified with MS										
Type = protein functional category										
AAB: Amino Acid Biosynthesis				IS: Insertion Element						
CEM: Cell Envelope Membrane				Nuc: Nucleotides						
CIM: Central Intermediary Metabolism				PPM: Proteases and Protein Modifications						
CF: Cofactor Biosynthesis				TrL: Translation						
CP: Cellular Processes				TrR: Transcription and Regulation						
EM: Energy Metabolism				TrSP: Transport						
LM: Lipids				RR: Replication and Repair						
HP Hypothetical Protein				Unc: Uncategorized						
Uniquely Expressed Proteins Detected on P1 Gels										
S.sol#	IPG 3-10	Pct	IPG 4-7	Pct	IPG 6-11	Pct	MW	pI	Type	Function
SSO0109	12	48	—	—	—	—	26580	9.58	HP	Conserved hypothetical protein
SSO0216	24	53	—	—	—	—	48489	9.63	TRL	(tuF-1) Elongation factor 1-alpha
SSO0341	10	14	—	—	—	—	102979	7.12	TRL	(alaS) Alanyl-tRNA synthetase
SSO0558	24	43	—	—	—	—	66516	7.44	TRL	(metS) Methionyl-tRNA synthetase
SSO0641*	51	18	47	11	—	—	118204	5.50	AAB	(carB) Carbamoyl-phosphate synthase large subunit
SSO0662	10	49	—	—	—	—	26757	5.08	CF	(bioR) Biotin-(acetyl-CoA carboxylase) synthetase
SSO0723	10	39	—	—	—	—	37089	6.34	AAB	(leuB) 3-isopropylmalate dehydrogenase
SSO0738	8	44	—	—	—	—	26588	5.82	PPM	Proteasome subunit
SSO1088	7	45	58	45	—	—	23232	4.74	HP	Conserved hypothetical protein
SSO1152	16	51	—	—	—	—	46122	9.50	HP	Conserved hypothetical protein
SSO1159	8	48	—	—	—	—	23009	7.52	Unc	Haloacetate dehalogenase
SSO1322	10	37	—	—	—	—	37185	5.44	AAB	(ilvC-2) Ketol-acid reductoisomerase
SSO1755	7	40	—	—	—	—	25160	5.69	HP	Conserved hypothetical protein
SSO1817	19	52	120	52	—	—	33655	6.90	CIM	(cysA-2) Thiosulfate sulfurtransferase
SSO2486	15	35	—	—	—	—	63230	7.03	TRL	(thrS) Threonyl-tRNA synthetase
SSO2585	—	—	—	—	83	64	36394	8.86	EM	(sqdB) L-lactate dehydrogenase
SSO2762	11	59	—	—	—	—	30620	10.15	EM	(etfA) Electron transfer flavoprotein, subunit alpha
SSO2766	21	55	—	—	—	—	52041	6.60	HP	Conserved hypothetical protein
SSO3187	11	47	—	—	—	—	33812	8.10	AAB	(serA-2) D-3-phosphoglycerate dehydrogenase
SSO3233	10	26	—	—	—	—	52327	8.32	CF	Cobyrinic acid synthase
* detected in tris-insoluble fraction										
Uniquely Expressed Proteins Detected on P2 Gels										
S.sol#	IPG 3-10	Pct	IPG 4-7	Pct	IPG 6-11	Pct	MW	pI	Type	Function
SSO0581	—	—	—	—	57	24	30452	10.20	TrR	GTP binding protein, hypothetical
SSO1095	12/57	22/16	145	31	—	—	95774	6.58	EM	Aconitate hydratase
SSO1184	11/58	40/52	—	—	121	62	21251	9.64	CF	(entB-like1) Isochorismatase related protein
SSO1423	53	30	—	—	—	—	23104	10.02	HP	Hypothetical protein
SSO1662	12/81	28/32	62	20	—	—	63396	5.28	AAB	(huyB-1) N-methylhydantoinase B
SSO1663	27	43	49	25	—	—	76406	5.76	AAB	(huyA-1) N-methylhydantoinase A
SSO2122	—	—	57	13	—	—	55655	6.23	TRL	(gatA-3) Glutamyl-tRNA amidotransferase, SU A
SSO2401	—	—	47	21	—	—	26703	5.89	HP	Conserved hypothetical protein
SSO2420	36/61	40/20	—	—	88	50	68396	9.19	CP	AAA family ATPase
SSO2639	21/118	40/33	153	40	—	—	83013	6.35	EM	(cutA-4) Carbon monoxide dehydrogenase
SSO2819	—	—	—	—	51	34	43305	9.64	EM	(fixC) FIXC protein homolog
SSO3105	—	—	—	—	119	31	54841	6.60	PPM	Thermostable carboxypeptidase 1

Table 15.

Proteins Uniquely Expressed in P1 and P2, Their Biochemical Function and Pathway

SSO#: red = greater than 5-fold up-regulated and black = only detected in this strain

Uniquely Detected on P1 Gels, Grown in: 0.1% tryptone, 0.1% yeast extract, 0.1% sucrose

<u>S.sol#</u>	<u>Function</u>	<u>Biochemical Pathway</u>
SSO0109	Conserved hypothetical protein	unknown
SSO0216	(tuF-1) Elongation factor 1-alpha	involved in protein synthesis, formation of aminoacyl-tRNA-elongation factor 1-alpha-GTP
SSO0341	(alaS) Alanyl-tRNA synthetase, EC 6.1.1.7	protein synthesis
SSO0558	(metS) Methionyl-tRNA synthetase, EC 6.1.1.10	protein synthesis
SSO0641*	(carB) Carbamoyl-phosphate synthase large SU, EC 6.3.5.5	glutamate & pyrimidine metabolism, conversion of L-glutamine to L-glutamate, formation of carbamoyl phosphate
SSO0662	(bioR) Biotin--(acetyl-CoA carboxylase) synthetase, EC 6.3.4.15	biotin metabolism for cofactors and vitamins
SSO0723	(leuB) 3-isopropylmalate dehydrogenase, EC 1.1.1.85	protein synthesis
SSO0738	Proteasome subunit, EC 3.4.99.46	multicatalytic endopeptidase complex which acts on peptide bonds, protein turnover
SSO1088	Conserved hypothetical protein	unknown
SSO1152	Conserved hypothetical protein	unknown
SSO1159	Haloacetate dehalogenase, EC 3.8.1.3	forms glycolate which is used in glyoxylate/decarboxylate metabolism (intermediates for purine and TCA cycle)
SSO1322	(ilvC-2) Ketol-acid reductoisomerase, EC 1.1.1.86	protein synthesis
SSO1755	Conserved hypothetical protein	unknown
SSO1817	(cysA-2) Thiosulfate sulfurtransferase, EC 2.8.1.1	inorganic ion transport, thiosulfate + cyanide ----> sulfite + thiocyanate
SSO2486	(thrS) Threonyl-tRNA synthetase, EC 6.1.1.3	protein synthesis, detected an additional isoform of threonyl-tRNA synthetase
SSO2585	(sqdB) L-lactate dehydrogenase EC 1.1.1.27	anaerobic pyruvate metabolism, highly up-regulated with tryptone source
SSO2762	(etfA) Electron transfer flavoprotein, subunit alpha	electron transport chain intermediate
SSO2766	Conserved hypothetical protein	unknown
SSO3187	(serA-2) D-3-phosphoglycerate dehydrogenase, EC 1.1.1.95	biosynthesis pathway for glycine, serine, and threonine
SSO3233	Cobyrinic acid synthase	new enzyme for archaea, also found in Methanococcus, possible porphyrin and vit. B12 metabolism

Uniquely Detected on P2 Gels, Grown in: 0.1% tryptone

<u>S.sol#</u>	<u>Function</u>	<u>Biochemical Pathway</u>
SSO0581	GTP binding protein, hypothetical	unknown
SSO1095	Aconitate hydratase, EC 4.2.1.3	involved in TCA cycle and reductive carboxylic acid cycle (CO2 fixation), highly up-regulated
SSO1184	(entB-like1) Isochorismatase related protein, EC 3.3.2.1	produces chorismate, found in high molecular weight complex with SSO2420, tryptophan, ubiquinone biosynthesis
SSO1423	Hypothetical protein	unknown
SSO1662	(huyB-1) N-methylhydantoinase B, EC 3.5.2.14	leads to sarcosine, sarcosine --> glycine + formaldehyde + H2O2
SSO1663	(huyA-1) N-methylhydantoinase A	leads to sarcosine, sarcosine --> glycine + formaldehyde + H2O2
SSO2122	(gatA-3) Glutamyl-tRNA amidotransferase, SU A, EC 6.3.5.-	Glu-tRNA(Gln) amidotransferase, biosynthesis of glutamine
SSO2401	Conserved hypothetical protein	unknown
SSO2420	AAA family ATPase	protein turnover, chaperone, acts in assembly and disassembly of replication complexes, complexed with SSO1184
SSO2639	(cutA-4) Carbon monoxide dehydrogenase, EC 1.2.99.2	probable aldehyde oxidoreductase, but annotated as an aerobic carbon monoxide dehydrogenase
SSO2819	(fixC) FIXC protein homolog	homologous to electron transfer flavo-ubiquinone protein
SSO3105	Thermostable carboxypeptidase 1, EC 3.4.17.19	hypothetical protease, Zn metalloprotease family

It is interesting to note that the annotated CODH (SSO2639, large subunit, 83KDa) and N-methylhydantoinase subunits A and B were detected when P2 was cultured in tryptone as compared to P1 being cultured in a yeast/sucrose/tryptone mixture. Since we detected lower molecular weight CODHs in P1 (see Table 13), we investigated whether P1 could be similarly induced to express the larger CODH (SSO2639), as well as the N-methylhydantoinases, when cultured with only tryptone. P1 growth in tryptone is problematic since during growth the pH of the media increases to the point of cell lysis. However, we were able to grow P1 to an OD₆₅₀ of 0.22 (as opposed to 0.6 for P2) and perform 2-DE on the harvested cells. Figure 38 shows that when the microbe is carbon source limited with tryptone, CODH and N-methylhydantoinase are up-regulated.

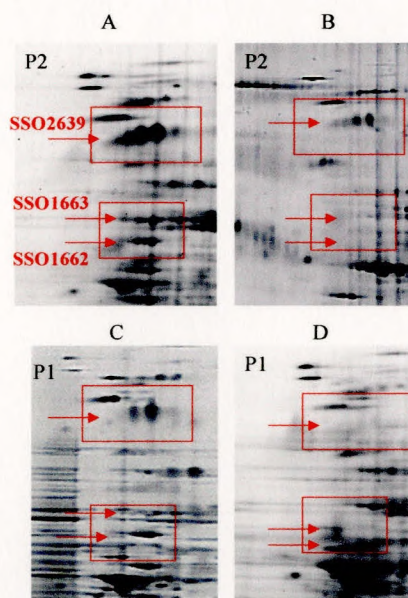


Figure 38. Induction of the large subunit carbon monoxide dehydrogenase (SSO2639) and N-methylhydantoinases (SSO1663 and SSO1662) in *S. solfataricus* P1 and P2 with the use of limited carbon sources. When P1 and P2 are cultured with 0.1% tryptone as the sole carbon and energy source (A and C), SSO2639, 1663, and 1662 are induced. When yeast extract and sucrose are added to the media (B and D), along with tryptone, little or no induction is detected.

Discussion

An examination of the P1 proteins identified indicate that many of the proteins participate in biochemical pathways associated with protein biosynthesis, as represented by three aminoacyl-tRNA synthetase proteins (alanyl, methionyl, and threonyl), the enzyme 3-isopropylmalate dehydrogenase, a keto-acid reductoisomerase, and 3-phosphoglycerate dehydrogenase (Table 15). Consistent with protein synthesis and turnover is the identification of an aminoacyl-tRNA elongation factor 1 α and a highly up-regulated endopeptidase complex (proteosome), respectively. Two protein involved in the metabolism of pyrimidines are also identified as evidenced by the tris-insoluble(aq) carbamoyl-phosphate synthases (SSO0641, phosphorylated and nonphosphorylated). Additionally, the production of glycolate from haloacetate dehydrogenase (SSO1159) may be providing glyoxylate to the tricarboxylic acid cycle, but the mechanism of conversion from glycolate to glyoxylate is unknown since the typical enzymes that fulfill this role (EC 1.1.1.26, 1.1.1.79, 1.1.1.29, 1.1.3.15, and 1.1.99.14) could not be identified from the annotation of the P2 genome.

Two proteins, biotin-acetyl-CoA carboxylase/synthase (SSO0662) and cobyrinic acid synthase (SSO3233) were surprising to find. SSO0662 obviously has some role in biotin metabolism, but it is unclear what types of cofactors and vitamins are being made. Menendez, et. al., suggested that *Sulfolobus metallicus* may utilize a novel autotrophic pathway, described as the 3-hydroxypropionate cycle, in which two molecules of CO₂ are fixed by biotin-dependent acetyl-CoA carboxylase and propionyl-CoA carboxylase and

glyoxylate is formed from acetyl-CoA (Menendez et al, 1999). It seems unlikely, however, that CO₂ fixation would be occurring for *S. solfataricus* P1 given the ample amount of carbon mixture being supplied (0.1% tryptone, 0.1% yeast extract, and 0.1% sucrose). It should be noted, however, that our carbon source mixture contained roughly thirty-times the amount of biotin in the yeast extract (6.6 ppb total) as compared to the amount of biotin typically identified in tryptone (Difco manual) and, hence, it is unknown how much this influenced the “up-regulation” of biotin-dependent acetyl-CoA carboxylase, as identified on the P1 gels. Cobyric acid synthase is a relatively new enzyme finding for archaea and was discovered in the cob operons of *Methanococcus* and *Pyrococcus* (Cameron et al, 1991; Eberhardt et al, 1997; Kyrpides et al, 2003). We believe this is the first evidence of cobyric acid synthase being expressed in archaea. Cobyric acid synthase is involved in porphyrin and vitamin B12 synthesis in bacteria (Kyrpides et al, 2003; Yan et al, 1997), and presumably its function in *Sulfolobus* is the same. There is a small amount of cyanocobalamine in yeast extract and tryptone powders (typically < 0.1 µg per g, Difco manual) and with a 0.2% stock it seems highly unlikely that the identification of cobyric acid synthase on our P1 gels had anything to do with this 0.2 ppb vitamin contamination in the stock powders.

Five proteins were identified in P1 that also have unknown functions (the conserved hypothetical proteins listed in Table 15). It seems reasonable that these hypothetical proteins may participate in protein biosynthesis and would be likely candidates to investigate if key enzymes are found missing in any of the typical protein biosynthesis pathways.

The differential regulation of proteins expressed by P2 when cultured with 0.1% tryptone as compared with P1 using a carbon mixture reflects the strict use of amino acids as the carbon and nitrogen sources (Table 15). A thermostable carboxypeptidase protein was up-regulated and is likely used to further degrade the polypeptides created from trypsin proteolysis of casein to produce additional amino acids for carbon and nitrogen utilization. Glutamyl-tRNA amidotransferase and isochorismatase, can be used to produce glutamine (Curnow et al, 1998) and chorismate (then ultimately tryptophan), respectively. Both P1 and P2 have been shown previously to produce heavy growth ($> 6 \times 10^8$ cell per mL) when glutamine is used as a nitrogen source or when a mixture of tryptophan and casamino acids is used as the carbon source (Grogan, 1989). Based on the composition of the yeast extract and tryptone stock used in this study (Difco manual), our carbon source mixture contained about twice the amount of total nitrogen and three times the amount of tryptophan compared to the tryptone-only media. It seems plausible, then, that if tryptone is used as the sole carbon and nitrogen source, this microbe may be preferentially assembling glutamine and tryptophan to promote growth. Interestingly, the isochorismatase protein was isolated along with a chaperone ATPase protein (SSO2420) in a high molecular weight complex on the basic side of 2-D gels and this complex was only visible when 0.1% tryptone was used (Figures 36 and 37). The type of ATPase identified (AAA+ ATPase) is known to act in the assembly and disassembly of replication complexes (Ruepp et al, 2000), but how it is functioning with isochorismatase in *Sulfolobus* is unknown.

The annotated carbon monoxide dehydrogenase (SSO2639, large subunit) identified in our experiments shares sequence homology to the N-terminally sequenced aldehyde oxidoreductase isolated from *Sulfolobus acidocaldarius* by Kardinahl, et al., (1999). In that study, the aldehyde oxidoreductase was shown to have oxidizing activity toward glyceraldehyde, glyceraldehyde 3-phosphate, isobutyraldehyde, formaldehyde, acetaldehyde, and propionaldehyde, using 2,6-dichlorophenolindophenol (Cl₂Ind) as the artificial electron acceptor. Further, it was shown to contain one molybdate, a FAD cofactor, one molybdopterin guanine dinucleotide, and two 2Fe-2S clusters. Growth of *S. acidocaldarius* using glucose as the sole carbon and energy source was found to be strictly molybdate-dependent in order to achieve high cell densities (OD₆₀₀ of 1.2) and revealed the presence of the reddish aldehyde oxidoreductase when molybdate (30 µg, Na₂MoO₄•H₂O) was added (Kardinahl et al, 1999). Without the supplemented molybdate, cell densities could only reach an OD₆₀₀ of 0.6. Thus, Kardinahl et. al., suggests that the holo-aldehyde oxidoreductase enhances the degradation of glucose by catalyzing the conversion of glyceraldehyde to glycerate in the nonphosphorylating Entner-Doudoroff (glycolysis) pathway and that this is reason behind the increased growth yields. Growth on glutamate or casein hydroslyate (tryptone), on the other hand, was independent of molybdate, routinely growing to an OD₆₀₀ of 1.2. The authors were unclear as to whether or not the aldehyde oxidoreductase was present in the cultures containing the amino acids, with and without molybdate.

In our study, we found the expression of the aldehyde oxidoreductase (SSO2639) in the presence 15 µg molybdate (Na₂MoO₄•H₂O) to be down-regulated or not detectable

when sucrose was included to the growth media and highly up-regulated when the carbon source was limited to 0.1% amino acids, for *S. solfataricus* P1 and P2 (Figure 38). We also noticed that expression of the N-methylhydantoinase subunits A and B (SSO1663 and SSO1662, respectively) mimicked the expression of the aldehyde oxidoreductase (SSO2639). Since N-methylhydantoinase is an essential enzyme in the conversion of creatine to sarcosine, and sarcosine can be converted to glycine, formaldehyde and H_2O_2 by sarcosine oxidase, we speculate that the holo-aldehyde oxidoreductase may be detoxifying the cell by oxidizing formaldehyde to formic acid and perhaps also be participating in an energy conservation pathway by funneling the transferred electrons down a branch of the electron transport chain of *S. solfataricus*. We believe that the conversion of creatine to sarcosine rather than to creatine phosphate probably indicates an energy deficient cellular state. The activity of the holo-enzyme has already been demonstrated by Kardinahl, et. al., and the added up-regulation of FIXC, an electron transfer flavoprotein-quinone oxidoreductase, with both the aldehyde oxidoreductase and N-methylhydantoinases further suggests the transfer of electrons to the respiratory chain.

The annotated genome of *S. solfataricus* P2 (She et al, 2001) does not contain any formaldehyde dehydrogenase (EC 1.2.1.1), however, it does encode 13 different types of aerobic carbon monoxide dehydrogenases (CODH), including SSO2639. These are not the same as the CODHs used by anaerobic organisms in the reductive acetyl-CoA cycle. In our extensive proteomic mapping of P2 in Chapter 5, we identified 8 out of 13 potential aerobic CODHs being expressed in tryptone media, including three large subunit types: SSO2639, SSO3009, and SSO3239. When P1 was grown with a mixture

of carbon sources that included sucrose, both SSO3009 and SSO3239 were again present on the 2-D gel, but SSO2639 was completely absent (Table 13).

Aerobic CODH has been identified and characterized in a large number of carboxydophilic bacteria, including: *Oligotropha carboxydovorans*, *Bradyrhizobium japonicum*, *Pseudomonas carboxydovorans*, *Streptomyces thermoautotrophicus*, and *Bacillus schlegelii* (Lorite et al, 2000; Meyer et al, 1986; Meyer et al, 1990; Meyer et al, 2000; Schubel et al, 1995). Aerobic CODH mediates the oxidation of CO to CO₂, using water as the oxidant: $\text{CO} + \text{H}_2\text{O} + \text{X(ox)} \rightarrow \text{CO}_2 + \text{XH}_2(\text{red})$ (Meyer et al, 1986). The anabolic function of CODH is to provide CO₂ for fixation and to funnel electrons down specific branches of the respiratory chain (Meyer et al, 1986; Meyer et al, 1990). The CODHs from different bacteria are composed of several subunits: 70-85 KDa (large), 25-33 KDa (medium), and 14-17 KDa (small). Complexes of L₃M₃S₃ and L₂M₂S₂ have been isolated (Meyer et al, 1986). We used a variety of detergents and a reducing agent and, thus, only identified the individual subunits (Barry et al, 2003b). Bacterial CODH's are also known to be novel molybdenum-containing iron-sulfur flavoproteins (Meyer et al, 1986), which was also true of Kardinahl's aldehyde oxidoreductase, but it remains to be seen what type of cofactors are assembled in SSO2639 of *S. solfataricus*. Additional study will also be required to determine if there are any stimulatory effects on the respiratory chain of *Sulfolobus* which are similar to those found in carboxydophiles. Carboxydophilic bacteria contain an electron transport system with cytochrome a-, b-, and c-type, while *Sulfolobus* combines homologs of cytochrome b and heme-a bearing cytochrome a's into SOX complexes (Schafer et al, 1996; Iwaski et al, 1995). It remains

to be determined if *S. solfataricus* contains electron transport chains similar to either *S. acidocaldarius* (which contains a SOX-ABCD and SOX-M components) (Shafer, 1996) or to *S. tokodaii* (Iwaski et al, 1995) and the type(s) of cytochrome(s) which act as electron acceptors.

Finally, there is the question about how SSO2639 should be annotated. SSO2639 shares strong sequence similarity to the large subunit CODH of carboxydophilic bacteria, including *Rhizobium loti* (E-score of e-117, with 33% identity at 261/778, 49% positives with 386/778 matches, and 4% gaps of 34/778), *Oligotropha carboxydovorans*, *Bradyrhizobium japonicum*, and *Pseudomonas carboxydoflava*, to name a few. However, it is unlikely that CO is present in our culture flasks and so the putative CODHs must be functioning in reactions which do not require CO. As mentioned earlier, SSO2639 (83.0 KDa) also matches the N-terminal sequence of the large subunit protein of Kardinahl's oxidoreductase (80.5 KDa) with 52% identity (13/25) and 76% positives (19/25). However, SSO3239 (79.0 KDa) also matches to the N-terminus of Kardinahl's proteins with 50% identity (13/26) and 65% positives (17/26). But only the expression of SSO2639 was dramatically affected by the carbon source used. Thus, it is possible that various annotated large, medium, and small subunit aerobic CODHs identified in *S. solfataricus* function in a variety ways, but most likely as aldehyde oxidoreductases and electron transport proteins. It also remains to be determined if any species of *Sulfolobus* is capable of chemolithoautotrophic growth with CO₂ or H₂ plus CO₂, and if it can it would be the first carboxydophilic archaeon identified, to our knowledge.

Concluding Remarks

The identification of *S. solfataricus* P1 proteins using the published *S. solfataricus* P2 genome database has now been established and allows for comparative studies between the strains. It may also be possible that proteins from the strains isolated by DeRosa (MT3 and MT4, de Rosa et al, 1975) from the same habitat as P1 and P2 could be identified using the P2 genome database. A total of 420 protein spots (224 gene products) from 2-D gels of P1 were mapped and identified using PMF and the P2 database.

Further by using the P2 database, we demonstrated that proteomics could be used to rationalize the physiological responses of *S. solfataricus* to changes in carbon sources. Additional applications of this type might be used to explain the idiosyncratic growth characteristics typically observed for different strains of *Sulfolobus*. Difference gel analysis identified 28 proteins uniquely expressed by P1 when a carbon source mixture of 0.1% tryptone, 0.1% yeast extract, and 0.1% sucrose was used as compared to 14 unique spots on P2 gels when only 0.1% tryptone was used. Analysis of the detected P1 proteins indicated that many were implicated in protein synthesis and, surprisingly, the up-regulation of biotin-acetyl-CoA carboxylase/synthase and cobyric acid synthase. The role of these enzymes in *Sulfolobus* metabolism has yet to be determined. Further, we believe this is the first example of an archaeon expressing the cobyric acid synthase protein, although it had been identified in the genome of *Methanococcus* and *Pyrococcus*.

Identification of the uniquely detected spots for P2 identified a potentially new detoxifying pathway for *S. solfataricus*, which may also act in an energy conservation pathway. Identification of the cutA-type carbon monoxide dehydrogenase (CODH, SSO2639), along with N-methylhydantoinases (SSO1662 and SSO1663) suggests that SSO2639 may be up-regulated to detoxify the production of formaldehyde from the degradation of creatine. Thus, we believe that SSO2639 is more likely an aldehyde oxidoreductase than an aerobic CODH. Presumably the source of the creatine is from the production of the metabolism of arginine in the urea cycle, which may be stimulated by using amino acids as the sole nitrogen source and correlates with the increased production of glutamine. Further, we also suspect that P2 and P1, when grown on 0.1% tryptone, were in an energy conservation mode and most likely funneling electrons to a specific branch of the electron transport chain. Energy conservation and electron transfer are supported by the known function of similar FAD containing aerobic CODHs and the up-regulation of FIXC and tryptophan, an electron transfer flavoprotein-quinone and a precursor to nicotinamide products, respectively.

Further study will be need to confirm whether SSO2639 is an aldehyde oxidoreductase or not. With the large number of annotated CODHs in *S. solfataricus* P2 (as well as in *S. tokodaii*), it remains to be seen what the true function is for each of them. Since many key metabolic enzymes are currently missing or unknown, it may turn out that some of these CODHs function as oxidoreductases in many different metabolic pathways. It will also be interesting to determine if *Sulfolobus* is at all carboxydrotrophic. Given that oxygen solubility is low at high temperatures and that *Sulfolobus* can be found

in microbial mats (from which CO producing bacteria may inhabit), as well as the habitat similarities between this hyperthermophilic archaeon and the thermophilic carboxydophilic bacterium, *Streptomyces thermoautotrophicus* (isolated from burning charcoal piles) it is possible that *Sulfolobus* could turn out to be the first carboxydophilic archaeon identified.

Overall, we have demonstrated a technique that can now be used to study the differences between the various *Sulfolobus* strains (Fusi et al, 1993; Martusewitsch et al, 2000; Prangishvili et al, 2000; Sehgal et al, 2002). We have also identified several proteins which suggest that *Sulfolobus* may use novel biochemical pathways which are as yet unknown or poorly understood, and we have identified several candidate proteins which may be used to help decipher those pathways.

CHAPTER 7

PROTEOMIC RESPONSE MECHANISMS OF EXTREMOPHILIC
ARCHAEA TO ARSENITE AND ARSENATEIntroduction

MSU has recently established the Thermal Biology Institute to study the organisms in the thermal environments of Yellowstone National Park (www.tbi.montana.edu). The ability to cope with harsh environmental variables, like high temperature and high levels of toxic metals, presumably requires the existence of unique cellular machinery. We used proteomic techniques in this study to display those proteins used by *Sulfolobus* to cope with high levels of arsenite, As(III), or arsenate, As(V). The proteins regulated in response to the different levels of arsenite and arsenate were determined. Since our previous studies clearly demonstrated the need for analytical rigor when comparing protein expression on 2-D gels (Chapter 4), we judged it important to perform all of our experiments in triplicate. That meant running three control gels and comparing the information from those gels (as a group) with three gels representing the stressed condition. We developed a differential gel analysis technique using whole cell extracts, rather than aqueous tris-soluble and aqueous tris-insoluble subfractionation, to reduce the number of gels required to monitor the proteome, while still performing all experiments in triplicate. Based on the information from this study, additional experiments are currently underway to try and elucidate the mechanisms of the arsenic

stress-response, as partially uncovered in this research. This is the first study of its kind that uses proteomics to discover the proteins regulated in archaea in response to two different concentrations of two different arsenous species. Further, since toxic metal resistance proteins have yet to be discovered in any archaea, the proteins which are determined to be key players in resisting arsenic in *Sulfolobus* may be used to illuminate the arsenic resistance system used by some archaea.

Bacterial and eucaryal resistance to toxic metals has been well studied. (Arsenic is often referred to as a toxic metal, but technically it is a metalloid. For the sake of continuity with previous studies of arsenic, arsenic will be referred to as a metal.) Bacterial genomes and plasmids have been shown to encode resistance systems for a variety of toxic metal ions, including Ag^+ , AsO_2^- , Cd_2^+ , Co_2^+ , CrO_4^{2-} , Cu^{2+} , Hg^{2+} , Ni^{2+} , Pb^{2+} , Sb^{3+} , TeO_3^{2-} , Tl^+ , and Zn^{2+} (Silver, 1996). Generally, resistance is accomplished in bacteria by enzymatic reduction and/or chemiosmotic excretion via unique membrane protein pumps (Silver, 1996). Eucaryotic cells seem to resist metal toxicity by inducing the expression of stress proteins. Toxic metal resistance systems of archaea, on the other hand, have not received much attention, as yet.

Geothermal springs within YNP often contain arsenic at concentrations of 10-40 μM , levels considered toxic to many organisms (Langner et al, 2001). Arsenic is one of the most environmentally troublesome genotoxicants known and appears in several inorganic and organic species in the environment. Arsenite, As(III) , seems to be the most toxic form and usually appears in biological systems as the uncharged arsenous acid H_3AsO_3 ($\text{pK}_a=9.2$) (Gebel, 2001). Arsenate, As(V) , is charged in biological media,

found as arsenate anions H_2AsO_4^- and HAsO_4^{2-} ($\text{pK}_a = 7$), and is not considered genotoxic (Gebel, 2001). Studies performed at MSU showed that arsenite is the most predominant valence state at the point of discharge in YNP hot-springs, but gets rapidly oxidized downstream to arsenate by organisms living in Fe/As-rich material (Langner et al, 2001).

Some bacteria are known to contain a resistance system that reduces As(V) to As(III) enzymatically with the protein ArsC and then pumps the arsenite out of the cell using the membrane protein ArsB. ArsB can function alone or in conjunction with the protein ArsA, as an ArsA-ArsB ATPase (Silver, 1996). Eucaryotic cells seem to resist arsenic toxicity by inducing the expression of heat-shock proteins (the types of which depend on the valence state of arsenic among other factors) with arsenite being the most potent inducer of heat shock proteins in organs and cells (Del Razo et al, 2001).

Recently, Banfield and Gihring compiled all of the known arsenic resistance proteins from genes and proteins which were available in the public databases and found evidence that the gene which encodes the arsenite efflux protein, known as ArsB, showed a parallel evolutionary path to the 16S rDNA sequences between archaea and bacteria (Gihring et al, 2003). We have also found evidence that such systems might exist in the extremophilic archaeon, *Sulfolobus*. Thus, the arsenite efflux protein pump may be present in a common ancestor(s) between bacteria and archaea. Blastp searches (which perform amino acid sequence homology comparisons) between the amino acid sequences of bacterial toxic metal resistance proteins against the genome of *Sulfolobus solfataricus* P2 (http://www-archbac.u-psud.fr/cgis/Pab_r/Blast2Server.acg) revealed the presence of highly homologous sequences in *Sulfolobus*. Table 16 lists Blastp results and

the putative *Sulfolobus* proteins found (along with their corresponding pI and MW). In particular, *Sulfolobus* was found to encode putative proteins with very high sequence homology to proteins involved in the enzymatic reduction of mercury, namely merA and merP; moderately matched to an arsenic regulatory protein from a member of the ArsR family, ArsR; and weakly matched to an arsenite efflux pump, ArsB. The lack of a significant E-value for the presence of ArsA suggests that if *Sulfolobus* does possess the ability to reduce arsenate, it probably pumps it out chemiosmotically rather than using an ArsA-ArsB ATPase. These Blast results are interesting and further study should aid in understanding the biology and the phylogenetic relationships of heavy metal detoxification. To date, no one has isolated toxic metal resistance proteins in archaea.

Table 16. <u>Blastp results</u>						
Known resistance system proteins homologous to ORFs of <i>S. solfataricus</i> P2						
Protein	Score	E-value	Identities	Positives	<i>Sulfolobus</i> Prot.	pI/MW ^a
MerA ¹	225	6.00E-60	148/448	230/448	SS02689	6.81/49.4
MerP ¹	48	1.00E-07	23/58	33/58	SS02896	8.43/76.3
ArsR ²	53	5.00E-09	28/66	41/66	SS02688	8.45/12.8
ArsB ²	39	8.00E-04	70/352	139/352	SS00144	8.79/45.5
^a pI represents isoelectric point and molecular weight (MW) is in KDa. pI and MW are for <i>Sulfolobus</i> proteins. ¹ Protein sequences were from plasmid 100 of <i>E. coli</i> K12. ² Ars protein sequences were from the plasmid R46 of <i>Salmonella typhimurium</i> .						

Electrophoresis Materials and Equipment

All materials and equipment was as described in Chapter 5. P2 was cultured as described in Chapter 5, but were supplemented with 0.1% tryptone, 0.1% yeast extract, 0.1% sucrose and 250 μ M inorganic phosphate.

Method Development

In order to monitor the proteomic response mechanisms of *S. solfataricus* P2, we cultured P2 in the presence of 350 μ M and 1 mM arsenite and compared the protein expression levels with those of P2 cultures without arsenite. We performed the same comparisons using 350 μ M and 1 mM arsenate. In all cases we found it necessary to include 250 μ M Phosphate in the medium to help *Sulfolobus* fight the arsenic toxicity and still grow. Protein extracts were prepared according to the methods used in Chapter 4, materials and methods. Proteins from whole cell extracts, solubilized in 5M urea, 1.4M thiourea, 3%Chaps,1% ASB-14, 4mM Tributyl Phosphine, and 0.2% Carrier Ampholytes, were separated with two-dimensional gel electrophoresis (2-DE), using over-lapping immobilized pH gradients (IPG) isoelectric focusing and nonlinear 8-18% gradient polyacrylamide gels (SDS-PAGE). All runs were performed in triplicate and composite gels from the control were compared with composite gels from the arsenite stress conditions. PDQuest image analysis software was used to highlight those proteins which were differentially expressed by the student t-test.

Results and Discussion

Gel image analysis results, which compare the control to the stressed cells, are displayed in Figures 39 and 40. The images are representations of composites from three control gels and three stress gels (see Chapter 4 for more detail) created using the software and illustrates how the PDQuest program performs the quantitative evaluations of the protein spots.

Arsenite Exposure

When exposed to 350 μ M arsenite, 127 protein spots were identified as differentially expressed on IPG 4-7 gels; while 33 spots were identified on the IPG 6-11 gels. When exposed to 1 mM arsenite, 84 proteins were highlighted on IPG 4-7 gels, while 30 were shown to be differentially expressed in the IPG 6-11 gels.

Arsenate Exposure

Exposure to 350 μ M arsenate resulted in 63 protein spots being identified as differentially expressed; while 50 were found in the IPG 6-11 gels. Exposure to 1 mM arsenate resulted in 84 differentially expressed spots on the IPG 4-7 gels, while the IPG 6-11 gels displayed 113 spots.

Spot Harvesting and Spot Identification

All of the spots differentially expressed spots were plucked, digested with trypsin, analyzed with MALDI MS, and subjected to database matching. A total of 584 spots were identified as differentially expressed proteins and identified using PMF. In order to obtain a smaller list from which to work with (as an initial attempt to interpret the data), the list was narrowed down to reflect only those proteins which were regulated under two or more of the arsenic conditions used. It is believed that if a protein was regulated in response to arsenite or arsenate stress, then it should be affected in at least two of the four conditions investigated. There were, however, some proteins that were regulated only under one set of conditions and which have been previously identified by other researchers as being stress-response type proteins. Both the proteins regulated under at least two conditions and those additional proteins judged to be possibly significant are listed in Tables 17 and 18.

Concluding Remarks

Trying to interpret the results from the list of proteins in Tables 17 and 18 is a challenge. Arsenate is a phosphate analog and interferes with phosphate metabolism at many levels. Arsenite typically causes oxidative stress but can also react with thiols and methionine, as well as be oxidized to arsenate. At the time of this writing it is still too speculative to interpret the results obtained. Investigations are currently underway and

additional background research will have to be performed before metabolic explanations can be postulated. mRNA expression studies are currently being carried out on some of the proteins identified in this study by the McDermott lab at Montana State University. We were unable to find any evidence of ArsB or ArsR expression, and subsequent experiments to identify the presence of mRNA which transcribes ArsB have also failed.

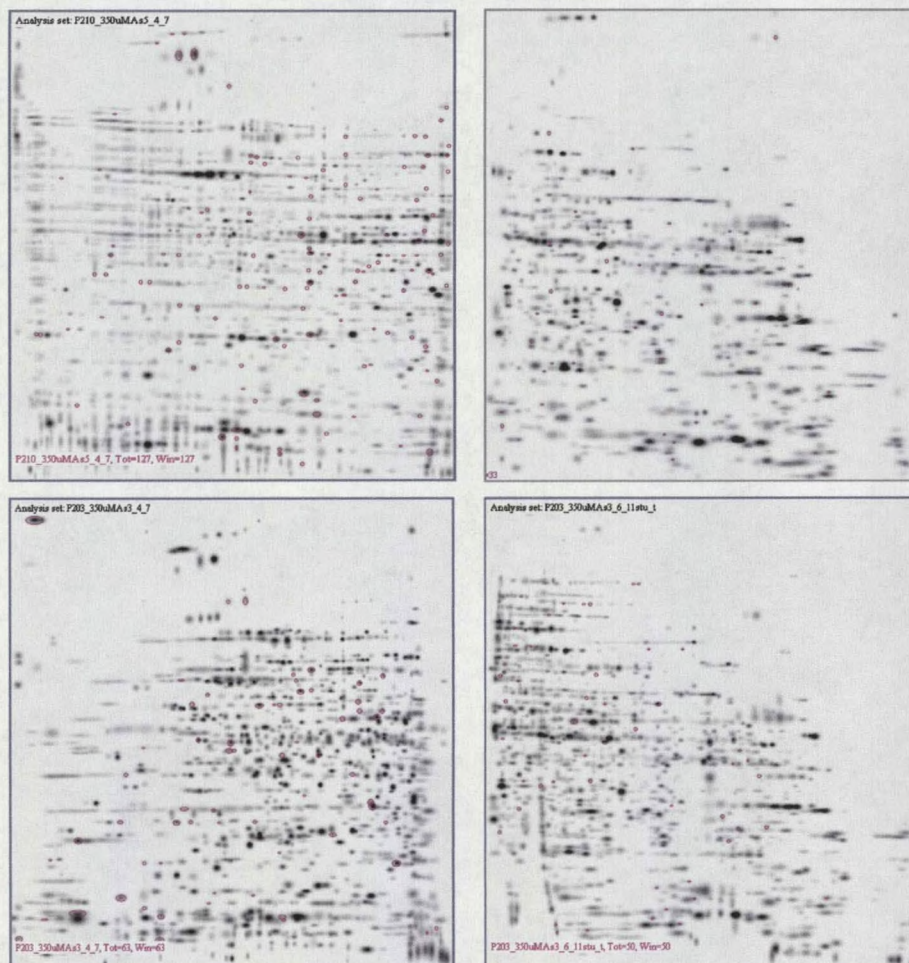


Figure 39. PDQuest image analysis results of 350 μ M arsenite exposure (top row) and 350 μ M arsenate (bottom row). Synthetic gels created from the triplicate control gels and triplicate stress gels. Spots determined to differ using a student's t-test (with a 95 % confidence interval) are encircled (purple color). IPG 4-7 gels (and 8-18% PAGE) are both to the left of the IPG 6-11 gels. The number of spots found differentially expressed are 127, 33, 63, and 50 (starting with the upper left image and going clockwise).

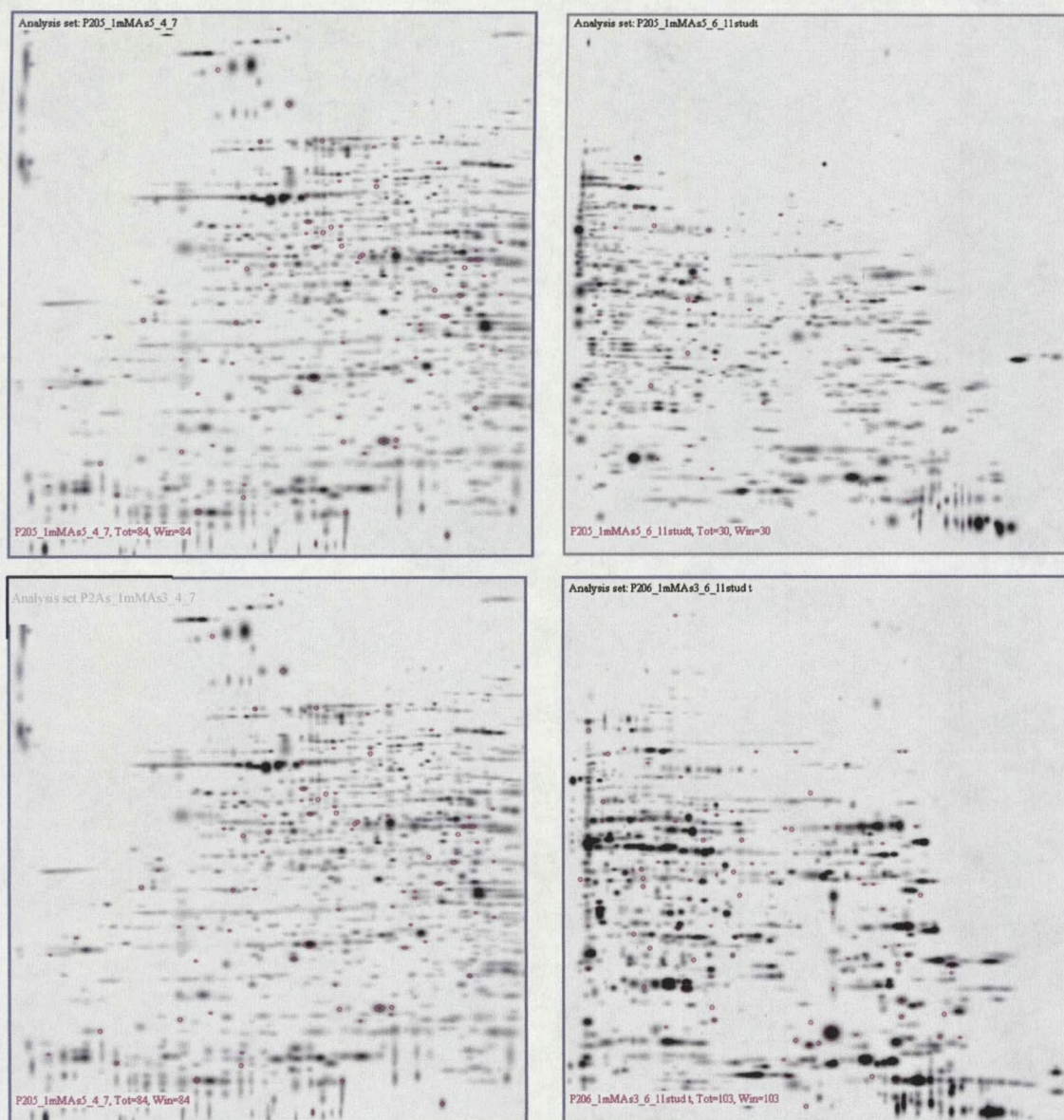


Figure 40. PDQuest image analysis results of 1 mM arsenite exposure (top row) and 1mM arsenate (bottom row). Synthetic gels created from the triplicate control gels and triplicate stress gels. Spots determined to differ using a student's t-test (with a 95 % confidence interval) are encircled (purple color). IPG 4-7 gels (and 8-18% PAGE) are both to the left of the IPG 6-11 gels. The number of spots found differentially expressed are 84, 30, 84, and 113 (starting with the upper left image and going clockwise).

Table 17 **Comparison of Proteins Regulated in More Than One Experiment**

Green: Only within As(III) or As(V)		Blue: down-regulated relative to P2 control			
Red: within As(III) and As(V)		Red: up-regulated relative to P2 control			
P2 ID	IPG 4-7	350uMAs5	1mMAs5	350uMAs3	1mMAs3
SSO0099	Conserved hypothetical protein	1.9	-	-	1.3
SSO0151	(hpS-1) D-arabino 3-hexulose 6-phosphate form	-	1.5	-	1.2
SSO0154	Hypothetical protein	-	3.5	-	1.6
SSO0282	(thsB) Thermosome beta subunit	-	3.3	1.7	2.0
				91.2	1.6
SSO0286	Conserved hypothetical protein	1.6	1.5	-	2.0
					8.3
SSO0397	(pcnA-like) Proliferating cell nuclear antigen putative	-	-	1.6	1.7
SSO0428	Hypothetical protein	-	88.4	-	6.7
SSO0527	Phosphoglycerate kinase (EC: 2.7.2.3)	1.5	-	-	1.5
SSO0530	(glyA) Serine hydroxymethyltransferase	58.3	-	1.7	-
SSO0551	Conserved hypothetical protein	8.8	-	-	4.2
SSO0576	(ilvC-1) Ketol-acid reductoisomerase	1.6	1.4	-	1.9
SSO1047	(pcnA-2) Proliferating cell nuclear antigen	-	10.6	125.9	-
SSO1088	Conserved hypothetical protein	1.3	1.4	-	-
SSO1098	Conserved hypothetical protein	14.8	-	-	-
		49.4	39.7	-	-
SSO1107	Hypothetical protein	5.2	1.7	3.4	7.1
SSO11196	Conserved hypothetical protein	-	8.7	-	17.5
SSO1126	Conserved hypothetical protein	2.5	-	-	39.7
SSO1208	(porG-1) Pyruvate synthase gamma chain	5.0	-	-	5.4
SSO1225	(todA) Toluene 1,2-dioxygenase system ferredoxin-	-	1.9	11.0	-
SSO1525	(pdhA-2) Pyruvate dehydrogenase, alpha subunit	-	-	-	50.9
SSO1526	(pdhB-2) Pyruvate dehydrogenase, beta subunit	4.5	3.6	-	-
SSO1662	(huyB-1) N-methylhydantoinase B	-	-	-	13.9
SSO1663	(huyA-1) N-methylhydantoinase A	11.8	-	-	-
SSO1668	(huyA-like) N-methylhydantoinase A related protein	-	-	-	14.4
SSO1907	(gdhA-2) NAD specific glutamate dehydrogenase	2.6	9.5	-	-
SSO1947	Hypothetical protein	4.7	1.6	119.2	57.1
SSO2044	(gdhA-4) NAD specific glutamate dehydrogenase	36.8	1.5	-	-
SSO2150	(cutA-3) Carbon monoxide dehydrogenase	2.4	1.3	3.3	6.2
SSO2182	(idh) Isocitrate dehydrogenase, probable	5.4	-	-	1.4
SSO2191	Conserved hypothetical protein	1.3	-	-	-
		1.3	-	1.2	-
SSO2280	(cysS) CysteinyI-tRNA synthetase	-	1.3	2.0	1.3
SSO2286	Hypothetical protein	1.3	2.0	-	-
SSO2339	Eukaryotic-type peptide chain release factor	-	3.4	-	3.3
SSO2356	(sdhA) Succinate dehydrogenase subunit A	-	2.7	30.2	2.3
			14.7	1.5	2.8
SSO2359	(sdhD) Succinate dehydrogenase subunit D	1.3	-	-	-
SSO2363	Conserved hypothetical protein	2.0	-	2.8	5.1
SSO2450	TATA binding protein (TBP)-interacting protein	106.5	-	-	26.4
SSO2479	Sugar phosphate nucleotidyl transferase	2.9	-	1.4	1.3
SSO2483	(sucC) Succinyl-CoA synthetase, beta subunit	7.9	-	171.8	-
SSO2632	Hypothetical protein	25.8	1.2	3.3	11.9
SSO2635	Hypothetical protein	3.0	-	4.1	14.6
SSO2636	(cutB-2) Carbon monoxide dehydrogenase	51.8	-	138.0	17.5
SSO2637	(cutC-2) Carbon monoxide dehydrogenase	127.8	-	-	-
SSO2642	(rr) Rubrerythrin	11.9	2.8	-	4.4
SSO2815	2-oxoacid--ferredoxin oxidoreductase, alpha chain	-	-	1.3	2.9
SSO2869	NAD-dependent malic enzyme	-	3.4	-	-
SSO2918	First ORF in transposon ISC1212	42.5	-	71.9	-

Table 17 **Other notable regulation:****Comparison of Proteins Regulated in More Than One Experiment**

Green: Only within As(III) or As(V)

Blue: down-regulated relative to P2 control

Red: within As(III) and As(V)

Red: up-regulated relative to P2 control

P2 ID	IPG 4-7	350uMas5	1mMas5	350uMas3	1mMas3
SSO0176	AAA family ATPase	-	2.7	-	-
SSO0250	(radA) DNA repair protein radA	-	26.4	-	-
SSO0271	AAA family ATPase	-	-	-	4.2
SSO0316	(sod) Superoxide dismutase [Fe]	-	19.0	-	-
SSO0421	AAA family ATPase	-	-	2.4	-
SSO0564	(atpB) ATP synthase subunit B	2.7	-	-	-
SSO0728	(EF-2) Elongation factor 2	8.4	-	-	-
SSO0738	Proteasome subunit (EC: 3.4.99.46)	-	1.3	-	-
SSO0755	(ahcY) S-adenosyl-L-homocysteine	4.2	-	-	-
SSO0758			9.3	-	-
SSO0876	(akH) Aspartokinase	-	3.4	-	-
SSO0959	DNA repair protein rad25	26.2	-	-	-
SSO0962	DNA binding protein SSO10b	176.2	-	-	-
SSO1045	Ribose phosphate pyrophosphokinase	14.5	-	-	-
SSO1195	(sixA) Phosphohistidine phosphatase	-	1.4	-	-
SSO1334	(aceB/mas) Malate synthase, putative	26.1	-	-	-
SSO1817	(cysA-2) Thiosulfate sulfurtransferase	-	2.3	-	-
SSO2156	Endonuclease IV related protein	-	47.4	-	-
SSO2205	(fabG-4) 3-oxoacyl-(acyl carrier protein)	9.6	-	-	-
SSO2427	Small heat shock protein hsp20 family	-	1.6	-	-
SSO2433	(cutC-1) Carbon monoxide dehydrogenase	-	1.4	-	-
SSO2493	(lipP-1) Lipase (EC: 3.1.1.3)	4.1	-	-	-
SSO2522	(php) Aryldialkylphosphatase	1.7	-	-	-
SSO2560	Nitroreductase, bluB homolog, putative	6.0	-	-	-
SSO2589	Citrate synthase	5.4	-	-	-
SSO2597	(agxI) Serine-pyruvate aminotransferase	5.4	-	-	-
SSO2639	(cutA-4) Carbon monoxide dehydrogenase	6.1	-	-	-
SSO2644	Hypothetical protein	69.7	-	-	-
SSO2737	Deacetylase, putative	6.0	-	-	-
SSO2817	(etfAB/fixAB) Electron transfer flavoprotein	-	-	-	30.8
SSO2909	(cysI) Sulfite reductase hemoprotein	-	-	-	6.3
SSO3000	(thsC) Thermosome gamma subunit	-	-	-	35.6
SSO3105	Thermostable carboxypeptidase 1	-	-	-	9.4
SSO3194	(gapN-3) Glyceraldehyde-3-phos dehydro	-	-	1.3	-
SSO3206	Protein synthesis inhibitor, putative	42.8	-	-	-
SSO6454	(snRNP-2) Small nuclear riboprotein protein	-	2.1	-	-

Table 18					
Comparison of Proteins Regulated in More Than One Experiment					
Green: Only within As(III) or As(V)					
Red: within As(III) and As(V)					
P2 ID	IPG 6-11	350uMAs5	1mMAs5	350uMAs3	1mMAs3
SSO0203	Conserved hypothetical protein	-	-	1.5	4.2
SSO0207	(pmM) Phosphomannomutase	-	46.0	2.5	-
SSO0209	Acetyltransferase, putative	-	-	4.2	35.9
SSO0216	(tuF-1) Elongation factor 1-alpha	2.3	-	-	10.7
SSO0241	Conserved hypothetical protein	1.8	-	-	1.4
SSO0316	(sod) Superoxide dismutase [Fe]	-	-	-	54.0
SSO0427	Protein L-isoaspartate methyltransferase	-	-	11.0	9.2
SSO0493	(thrC-1) Threonine synthase	-	-	4.3	1.5
SSO0561	(atpE) ATP synthase subunit E	2.3	-	3.5	-
SSO0579	(ilvB-2) Acetolactate synthase large subunit homolog	-	11.6	9.5	-
SSO0676	(moeA-1) Molybdopterin biosynthesis protein	2.9	-	-	5.2
SSO1107	Hypothetical protein	-	-	2.8	1.3
SSO11972	Conserved hypothetical protein	-	7.1	-	9.5
SSO1514	Conserved hypothetical protein	4.0	-	-	2.2
SSO2353	Geranylgeranyl hydrogenase, putative	-	-	1.4	13.7
SSO2364	(SSB) Single-stranded DNA binding protein	5.2	-	-	1.5
SSO2514	3-hydroxyacyl-CoA dehydrogenase	2.3	10.4	-	-
SSO2629	Oxidoreductase (flavoprotein)	-	-	2.2	-
SSO2639	(cutA-4) Carbon monoxide dehydrogenase	-	-	3.4	24.7
SSO2909	(cysI) Sulfite reductase hemoprotein	-	1.1	2.0	2.6
SSO3254	(rpoB2) DNA-directed RNA polymerase	-	10.4	3.4	-

Other notable regulation:					
SSO0098	Methionine aminopeptidase 2	-	-	-	2.6
SSO0421	AAA family ATPase	-	-	-	2.9
SSO0530	(glyA) Serine hydroxymethyltransferase	-	-	-	88.7
SSO0905	(serA-1) D-3-phosphoglycerate dehydrogenase	-	-	-	36.4
SSO0998	(nadA) Quinolinate synthetase	-	-	-	27.4
SSO1817	(cysA-2) Thiosulfate sulfurtransferase	-	-	2.5	-
SSO1960	Hypothetical protein	-	-	-	96.2
SSO2098	(tri) Tricorn protease	-	3.3	-	-
SSO2222	(trxB-1) Thioredoxin reductase	4.7	-	-	-
SSO2255	(bcp-3) Peroxiredoxin, bacterioferritin	-	-	-	4.2
SSO2356	(sdhA) Succinate dehydrogenase subunit A	-	6.4	-	-
SSO2357	(sdhB) Succinate dehydrogenase subunit B	2.2	-	-	-
SSO2357	(sdhB) Succinate dehydrogenase subunit B	-	-	1.4	-
SSO2416	(trxB-3) Thioredoxin reductase (EC: 1.6.4.5)	-	10.4	-	-
SSO2434	(cutB-1) Carbon monoxide dehydrogenase	-	-	-	4.9
SSO2912	(sat) Sulfate adenylyltransferase (EC: 2.7.7.4)	-	-	-	2.0
SSO3107	(ilvD) Dihydroxy-acid dehydratase (EC: 4.2.1.9)	-	15.7	-	-
SSO3194	(gapN-3) Glyceraldehyde-3-phos dehydro, NADP dep	-	7.0	-	-

CHAPTER 8

SUMMARY OF RESEARCH ACCOMPLISHMENTS

An integrated genomic and proteomic approach toward the study of complex biological systems is the way of the future. Genomic investigations can be used to explore the functions of genes and predict many functions of an organism, while proteomic studies can provide operational information on the gene products, identify missing links in key metabolic pathways, and supply a list of gene candidates which can be tested for function. Together, both of these techniques allow for an integrated view of the phylogeny, molecular biology, and physiology of an organism.

The focus of this research project was to develop and implement proteomic tools and show how they can be used study protein expression in extremophilic organisms. In the process it was hoped that discoveries would be made that would lay the foundation for more detailed studies in the future. As it turned out, all of my projects were quite productive and the volumes of data that were generated were staggering, over 50 gigabytes in total. The bottle-neck in this project, as in most proteomic projects, has been sorting through the data and interpreting the results. The volume of data is the single most troublesome issue that all research groups in this field are facing. Improved software is being written and software tools are being developed to help automate certain aspects of the techniques used, but much more work is needed. The cutting-edge right now is in so called "systems biology" where several groups are developing software that integrates the vast amounts of information in the literature on enzyme properties,

metabolic pathways and signaling networks to help decipher the significance of protein expression differences observed from proteomic studies. We plan to subject our arsenic results to two new programs that have just become available that suggest metabolic pathways involved with listed proteins and possible interactions between listed proteins (previous software which had reported the capability to do this were tried, but did not perform as advertised).

In applying proteomics techniques to the research described in this thesis (as well as to the other projects that I have worked on), I have strived to evaluate the reliability of the results. I believe that a rigorous analytical approach is important and believe that I have worked to higher standards than has been typical in this field. Statistical considerations are important in designing the methods, performing the experiments, and interpreting the results.

The information and results from this research study resulted in the publication of an extensive analytical and statistical evaluation of 2-D gel methods in the journal, *Electrophoresis*. Gel-to-gel reproducibility, spot matching efficiency, and the degree of spot quantity reproducibility were evaluated. Further, each of those parameters were assessed using four different sample loading methods and two different IEF ranges. The results of that study should demonstrate how to perform improved differential gel analysis. Methods which are used to evaluate expression differences should provide results that are statistically sound (which is not the case in most publications over the last 20 years). In addition, the study demonstrated that it is essential to use the cup-loading method when separating protein extracts on IPG 6-11 or other alkaline IEF ranges. In

addition, a simple and inexpensive assembly for cup-loading is diagramed and can be used in place of the expensive commercial apparatus on the Protean IEF cell (Bio-Rad).

This research produced two extensive mappings, along with protein identifications, of archaeal proteins from 2-D gels. A total of 867 proteins were mapped for *S. solfataricus* P2 and a total of 420 proteins mapped for *S. solfataricus* P1, both over three different IPG ranges. Most of the isoforms are thought to be due to post-translational modification. From the number of isoforms, it is evident that the average protein has about three post-translational modifications and thus, the "one gene, one protein" hypothesis is "null and void" for archaea. Some proteins have many isoforms and some have only one isoform detected. Admittedly a small fraction of the isoforms may have been created during sample handling, but the most common modifications during sample handling are amino acid oxidation and these do not produce isoforms that can be resolved on 2-DE. Thus, we believe that most of the post-translational modifications occurred *in vivo*. The vast number of PTMs present emphasizes the importance of using a proteomics approach toward studying physiological responses of organisms to different growth conditions (or stimuli). Additional studies using LC/MS/MS will be needed to identify the post-translational modifications and the locations of those modifications in the protein sequences in order to gain a better understanding of the regulation of those cellular pathways affected.

In the course of the 2-D proteomic mapping, a metabolic pathway which was known to be encoded by the genome, but had not been confirmed at the protein level, was identified for the first time. That *Sulfolobus* uses gluconeogenesis was in doubt but

evidence for this pathway was obtained by the identification of the phosphoenolpyruvate carboxykinase enzyme. Additional proteins were discovered, such as carbon monoxide dehydrogenase and geranylgeranyl synthase, which also suggest the use of metabolic pathways that are still unknown in *Sulfolobus*.

It has also been well known that *Sulfolobus* strains exhibit their own “idiosyncratic” growth curves depending on the mixture of inorganic ions, carbon sources, nitrogen sources, temperature and pH. Closely related strains have also been shown to have different susceptibilities to viral infection. Understanding the “cause and effect” of each of these variables has remained speculative because of the need for methods that can be used to illuminate the genetic response to those variables *and* the ability to apply the method to more than one strain of *Sulfolobus*. Since strains P1, P2, MT3 and MT4 have been demonstrated to be extremely similar, and because the P2 genome had been recently sequenced (April, 2001), it was determined whether proteins between closely related strains could be identified using the P2 database. This was a natural extension to how most researchers use bioinformatics, since their goal is usually to find proteins (or nucleic acids) which have homologous sequences or structures in order to predict the function of the protein being studied. It was demonstrated that the P2 database can be used to identify P1 proteins by mapping 420 spots on 2-D gels, which represented 224 homologous gene products. Using a proteomics approach, the proteome of each strain was used to examine the physiological state of the organism and to explain the growth characteristics observed for a particular strain. In the process, it was demonstrated that limitations on carbon and nitrogen sources will cause P1 and P2 to

express certain proteins which may indicate a type of cellular energy conservation.

Illumination of parts of this pathway suggest, for the first time, that *S. solfataricus* may be a carboxydotrophic archaeon or utilize a carbon fixation pathway that has yet to be characterized. Certainly, the discovery of the proteins identified will spark further research to investigate their meaning more thoroughly.

Differential gel analysis was applied to study the arsenic resistance of *S. solfataricus*. The results of that study have stimulated new research by the McDermott and Peters Laboratories, at MSU, to further study the meaning of the proteomic findings (at the biochemical and genomic levels). My study was the first to investigate the effect that different concentrations of arsenite and arsenate species had on archaea. The cells interconvert these arsenic species and we tested two concentrations of each. In current work we are striving to fit the matrix of protein regulation results to metabolic models, using some recently developed software tools.

Other accomplishments, that do not appear in this dissertation, include the development of two user guides: one for the rather tricky operation of automated MALDI MS software and the second for the operation of the very important image analysis software. Improper use of either method will results in poor or incorrect results and no one on the campus had previously learned to operate these complex software packages. The guides were written to give future proteomic researchers the benefit of my experience with these software packages and have been submitted to the makers of the software to be turned into application notes available to all users (they are currently available through Dr. Edward Dratz of the Department of Chemistry and Biochemistry, MSU). To further

encourage the use of proteomics, classical proteomic protocols were developed for three other research projects on campus: for the study of the development of the nervous system in chicken dorsal root ganglia, the control of axonal sprouting in the pituitary gland of the rat brain, and in the early development processes of xenopus frog eggs. The research tools and techniques put in place have applicability to a great many research areas. Many of the techniques are considered on the "cutting-edge" of science, and are the next big step in understanding biological complexity in the post-genomic world. I started my research hoping to make an impact in the world of science and to strengthening the research environment of MSU and I feel the techniques and discoveries made during my graduate study have done just that. At the time of this writing, I currently have one publication accepted by Electrophoresis, two first-author publications ready for peer review, submitted two different user manuals to be turned into application notes, in the process of writing another two first-author publications, and anticipate two second-authored collaborations (total 9). I hope that my work will continue to stimulate many new avenues for scientific discovery.

Appendix A

PROTEOMIC CHARACTERIZATION OF VIRAL INFECTION
IN ARCHAEA: APPLICATION TO *S. SOLFATARICUS*

PROTEOMIC CHARACTERIZATION OF VIRAL INFECTION
IN ARCHAEA: APPLICATION TO *S. SOLFATARICUS*

As an additional example of applying proteomic techniques toward studying the properties of *Sulfolobus solfataricus*, a study to identify the proteins regulated during viral infection was performed.

Ribosomal-RNA sequencing and other studies have established that many of these extremophiles represent a unique domain of life known as Archaea (formerly Archaeobacteria), distinct from Bacteria (eubacteria) and Eucarya (eucaryotes). Archaea share certain characteristics with both Bacteria and Eucarya. Among these common traits are the existence of viruses and plasmids. However, for certain types of Archaea, viruses and plasmids have been isolated and shown to contain very unique combinations of morphology and genome sequence. Many do not display sequence homology with any other known viral, archaeal, bacterial, or eucaryal genes. Investigations into the mechanisms employed by archaea to cope with extreme environmental variables and viral infections will provide information as to the cellular constituents utilized for adaptation and presumably will also illuminate the evolutionary relationships of these mechanisms.

Archaea can be subdivided into two kingdoms. 1) Euryarchaeota are a heterogeneous group of organisms that include the extreme halophiles and methanogens, the sulfur-dependent hyperthermophile *Thermococcales*, and the thermophilic/acidophilic Thermoplasmales (Granoff and Webster, 1999); 2) Crenarchaeota include the anaerobic

hyperthermophile *Thermoproteales* and the aerobic *Sulfolobales*, the latter being the subject of our investigations (Granoff and Webster, 1999).

Archaea (like bacteria and eucarya) have been found to host a variety of viruses (phages) and to contain plasmids. All but two phages of Euryarchaeota resemble bacteriophages. For example, extreme halophilic phages have the icosohedral head-to-tail morphology found in the lamda and T phages of *Escherichia coli* (Granoff and Webster, 1999). The best characterized halophage is Myoviridae (H) which contains temperate double-stranded DNA (dsDNA) that causes cell lysis when the host cells enter the stationary growth phase. A less common type of lysogeny occurs when the DNA is not inserted into the host genome but instead persists as an independently replicating DNA molecule in the host's cytoplasm.

In contrast to the viruses found in all the other kingdoms and domains of life, the viruses of Crenarchaeota appear to be unique in structure and genome sequence. These are the only known organisms that serve as hosts for filamentous (or rod-shaped) viruses containing linear dsDNA (Granoff and Webster, 1999) (Figure 41). Currently, no phages of lambdoid morphology (head-to-tail assembly) have been isolated from this kingdom. The best studied crenarchaeophages are the TTV and SSV viruses from *Thermoproteus* and *Sulfolobus*, respectively. For example, *Thermoproteus tenax* is an anaerobic hyperthermophile which serves as host for four different viruses: TTV1, TTV2, TTV3, and TTV4. These viruses differ in their DNA sequences and protein composition. TTV2 and TTV3 are filamentous while TTV1 is a flexible rod and TTV4 is a stiff rod (Granoff and Webster, 1999). *Thermoproteus* is a sulfur-dependent anaerobe which are difficult to

culture in the laboratory and to grow as lawns for plaque studies. Hence, investigations into the genome organization, protein content, and host-virus interactions of TTV has been difficult. The crenarchaeota *Sulfolobus*, on the other hand, can be routinely cultured and plaque tested.

Extrachromosomal elements of *Sulfolobus* are continuing to be discovered (insert reference to latest publication). Besides viruses, conjugative and nonconjugative plasmids have been isolated and characterized. There are currently four novel families of *Sulfolobus* viruses: *Fuselloviridae*, *Rudiviridae*, *Lipothrixviridae*, and *Guttaviridae* (Figure 36) (Prangishvili et al, 2001). *Fuselloviridae* viruses, SSV1, SSV2, and SSV3, are spindle (lemon)-shaped virions (60 x 100 nm) containing positively supercoiled circular DNA (15 Kbp) and an associated DNA-binding protein. The virion is enclosed within a hydrophobic coat protein envelope. The positive DNA supercoiling is surprising given that covalently closed circular DNA from viruses of bacteria and eucarya are, without exception, negatively supercoiled (Nadal et al, 1986). DNA sequence comparisons between the different SSVs show high sequence homology (Reiter et al, 1987). These viruses are temperate and appear to integrate into the Arg-tRNA gene of the host genome. Unlike temperate bacteriophages, *Sulfolobus* phages do not lyse or kill the host after induction. Instead, after viral production, a lysogenic state is maintained. The best-studied *Sulfolobus* virus is SSV1. It was originally discovered in the *S. shibatae* strain B12 and was shown in 1992 by the Zillig group to infect *S. solfataricus*, but not *S. acidocaldarius* (Schleper et al, 1992). SSV1 production can be induced by UV irradiation or mitomycin C treatment which is similar to induction of bacteriophages,

except that cell death does not occur (Prangishvili et al, 2001). This mechanism resembles UV induction of lysogenic lambdoid phages where UV light inactivates a host repressor of the short transcription unit, situated between two large back-to-back transcription units (Granoff and Webster, 1999).

The *Rudiviridae* viruses obtained from *S. islandicus*, SIRV1 and SIRV2, are stiff-rod shaped virions (90 x 850 nm) possessing superhelical linear dsDNA, a DNA binding protein, end plugs that seal in the DNA and tail filaments (Stedman et al, 1999). The lysogenic SIRVs are stable in their carrier state but differ dramatically in their genome stability. SIRV2 tends to conserve its sequence but SIRV1 accumulates point mutations at a rate of one every 1000 bases per replication cycle (Stedman et al, 1999). Interestingly, only *S. islandicus* strains could serve as hosts for the SIRVs and not *S. solfataricus*, *S. acidocaldarius*, or *S. shibatae*. However, SIRV1 could induce an SOS-type response in *S. solfataricus* and enhance the production of SSV1 virus, indicating that SIRV1 may enter *S. solfataricus*.

The two other families of *Sulfolobus* viruses, *Lipothrixviridae* and *Guttaviridae*, have only recently been discovered and are also lysogenic (Arnold et al, 1999; Arnold et al, 2000). *Lipothrixviridae* of *Sulfolobus islandicus* are flexible and filamentous virions (24 x 2000 nm) containing linear dsDNA and a zipper-like array of DNA-associated protein subunits (Arnold et al, 2000). *S. islandicus* seems to be the only host for this virus (*S. solfataricus* was not) and is named SIFV. The *Guttaviridae* virus, SNDV, was isolated in a *Sulfolobus* strain from New Zealand and is known as *Sulfolobus neozealandicus*, droplet-shaped virus. This virus is unique because no other virus has

been found containing cccDNA inside a bearded droplet-shaped virion. Morphologically this virion displays a surface that is helically ribbed. DNA sequencing shows evidence of dam-like methylation. Further, the SNDV genome is not integrated into the host chromosome and does not persist as a stable prophage, being lost upon prolonged incubation (Arnold et al, 2000).

Currently, there are very few studies investigating the virus-host relationship in archaea. This is surprising, given the dramatic genetic discoveries that have been made using bacteriophages. Crenarchaeophages may offer many of the same advantages for studying archaeal genetics as bacteriophages did for studying bacteria. Not only might these findings advance our knowledge of archaeal molecular genetics, but the findings could have evolutionary implications. Finally, infection with extremophilic viruses provides well-defined systems for developing and exercising proteomic methodology in archaea.

As an additional example of ongoing work, the classical proteomic techniques developed for differential gel analysis of arsenic stress were applied to study the virus-host interactions of *Sulfolobus solfataricus* P1 infected with SSV2. These preliminary results demonstrate yet another system and mechanism that can be studied. A number of proteins are up- and down-regulated after virus infection (Figure 41) and these proteins still need to be identified.

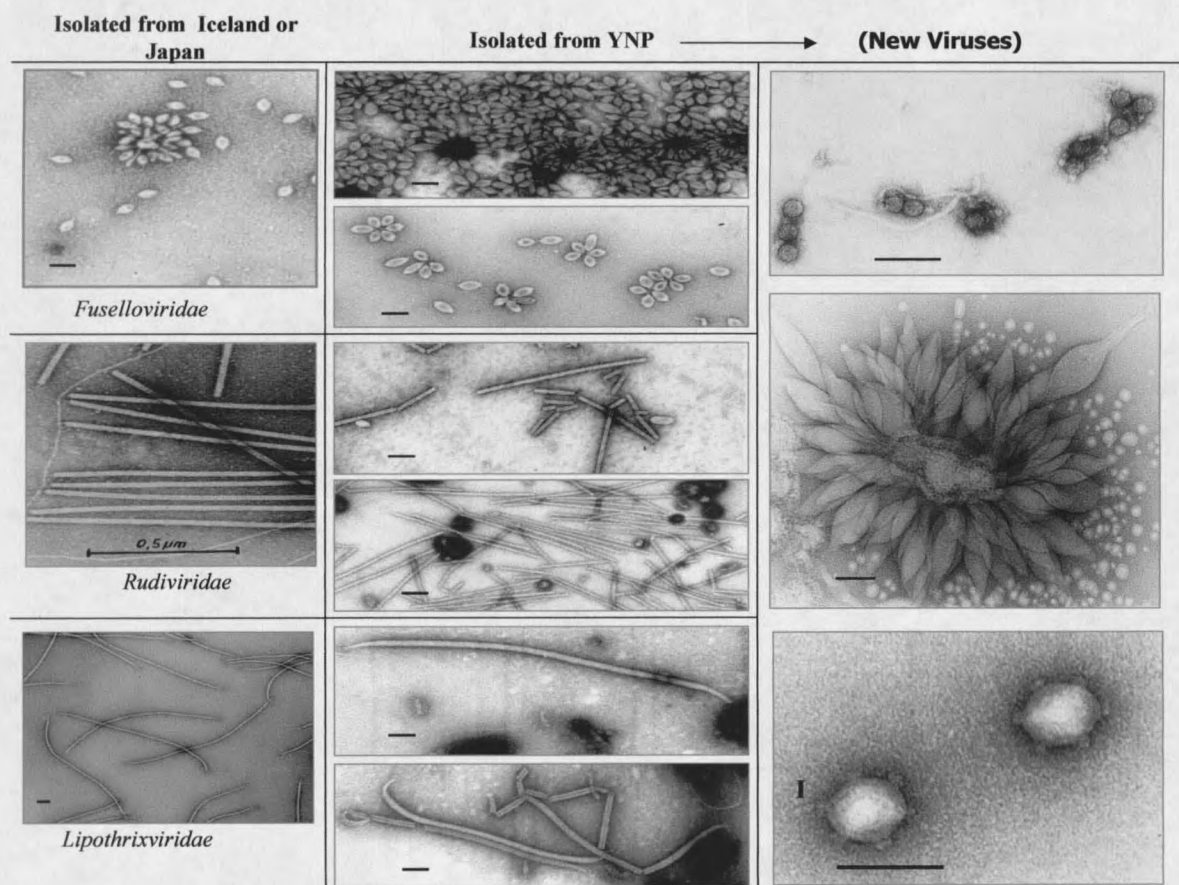


Figure 41. The viruses of *Sulfolobus*.



Figure 42. Response to SSV2 infection. 182 proteins were differentially expressed in response to SSV2 infection and isolated on the IPG 4-7 image (top) and 88 proteins were highlighted in the IPG 6-11 (bottom) image. Future work will seek to identify these proteins.

LITERATURE CITED

- Aagaard C, Leviev I, Aravalli RN, Forterre P, Prieur D, and Garrett RA (1996). General vectors for archaeal hyperthermophiles: strategies based on a mobile intron and a plasmid. *FEMS Microbiol Rev* 18:93-104.
- Aebersold R and Mann M (2003). Mass spectrometry-based proteomics. *Nature* 422:198-207.
- Adams, MW (1993). Enzymes and proteins from organisms that grow near and above 100°C. *Ann. Rev. Microbiol* 47: 627-658.
- Alaiya AA, Oppermann M, Langridge J, Roblick U, Egevad L, Brindstedt S, Hellstrom M, Linder S, Bergman T, Jornvall H, and Auer G (2001). Identification of proteins in human prostate tumor material by two-dimensional gel electrophoresis and mass spectrometry. *Cell Mol Life Sci* 58:307-311.
- Albers SV, van de Vossenberg JL, Driessen AJ, and Konings WN (2000). Adaptations of the archaeal cell membrane to heat stress. *Front Biosci* 5:D813-20.:D813-D820.
- Albers SV, van de Vossenberg JL, Driessen AJ, and Konings WN (2001). Bioenergetics and solute uptake under extreme conditions. *Extremophiles* 5:285-294.
- Andersen JS, Svensson B, and Roepstorff P (1996). Electrospray ionization and matrix assisted laser desorption/ionization mass spectrometry: powerful analytical tools in recombinant protein chemistry. *Nat Biotechnol* 14:449-457.
- Anderson L and Seilhamer J (1997). A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* 18:533-537.
- Aravalli RN and Garrett RA (1997). Shuttle vectors for hyperthermophilic archaea. *Extremophiles* 1:183-191.
- Arnold HP, She Q, Phan H, Stedman K, Prangishvili D, Holz I, Kristjansson JK, Garrett R, and Zillig W (1999). The genetic element pSSVx of the extremely thermophilic crenarchaeon *Sulfolobus* is a hybrid between a plasmid and a virus. *Mol Microbiol* 34:217-226.
- Arnold HP, Ziese U, and Zillig W (2000). SNDV, a novel virus of the extremely thermophilic and acidophilic archaeon *Sulfolobus*. *Virology* 272:409-416.
- Barns SM, Delwiche CF, Palmer JD, and Pace NR (1996). Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proc Natl Acad Sci, USA* 93:9188-9193.

- Barry RC, Alsaker BL, Robison-Cox JF, and Dratz EA. Quantitative evaluation of sample application methods for semi-preparative separations of basic proteins by two-dimensional gel electrophoresis. Electrophoresis. 2003a. In Press
- Barry RC, Young MJ, Stedman KM, and Dratz EA. Proteomic mapping of the extremophilic archaeon *Sulfolobus solfataricus* P2. Proteomics. 2003b. In Press
- Belov ME, Gorshkov MV, Udseth HR, Anderson GA, and Smith RD (2000). Zeptomole-sensitivity electrospray ionization--Fourier transform ion cyclotron resonance mass spectrometry of proteins. *Anal Chem* 72:2271-2279.
- Berggren K, Chernokalskaya E, Steinberg TH, Kemper C, Lopez MF, Diwu Z, Haugland RP, and Patton WF (2000). Background-free, high sensitivity staining of proteins in one- and two-dimensional sodium dodecyl sulfate-polyacrylamide gels using a luminescent ruthenium complex. *Electrophoresis* 21:2509-2521.
- Bjellqvist B, Ek K, Righetti PG, Gianazza E, Gorg A, Westermeier R, and Postel W (1982). Isoelectric focusing in immobilized pH gradients: principle, methodology and some applications. *J Biochem Biophys Methods* 6:317-339.
- Bohm G and Jaenicke R (1994). Relevance of sequence statistics for the properties of extremophilic proteins. *Int J Pept Protein Res* 43:97-106.
- Brock T.D., Brock K.M., Belly R.T., and Weiss RL (1972). *Sulfolobus*: A new genus of sulfur-oxidizing bacteria living at low pH and high temperature. *Arch Mikrobiol* 84:54-68.
- Brush M (1998). Dye Hard: Protein Gel Staining Products. *The Scientist* 12:16-22.
- Burke TW, Mant CT, Black JA, and Hodges RS (1989). Strong cation-exchange high-performance liquid chromatography of peptides. Effect of non-specific hydrophobic interactions and linearization of peptide retention behaviour. *J Chromatogr* 476:377-89.:377-389.
- Burlingame AL, Millington DS, Norwood DL, and Russell DH (1990). Mass spectrometry. *Anal Chem* 62:268R-303R.
- Cameron B, Guilhot C, Blanche F, Cauchois L, Rouyez MC, Rigault S, Levy-Schil S, and Crouzet J (1991). Genetic and sequence analyses of a *Pseudomonas denitrificans* DNA fragment containing two cob genes. *J Bacteriol* 173:6058-6065.
- Carr SA, Hemling ME, Bean MF, and Roberts GD (1991). Integration of mass spectrometry in analytical biotechnology. *Anal Chem* 63:2802-2824.

- Chait BT and Kent SB (1992). Weighing naked proteins: practical, high-accuracy mass measurement of peptides and proteins. *Science* 257:1885-1894.
- Cheon MS, Fountoulakis M, Dierssen M, Ferreres JC, and Lubec G (2001). Expression profiles of proteins in fetal brain with Down syndrome. *J Neural Transm Suppl* 311-319.
- Chevallet M, Santoni V, Poinas A, Rouquie D, Fuchs A, Kieffer S, Rossignol M, Lunardi J, Garin J, and Rabilloud T (1998). New zwitterionic detergents improve the analysis of membrane proteins by two-dimensional electrophoresis. *Electrophoresis* 19:1901-1909.
- Cohen GN, Barbe V, Flament D, Galperin M, Heilig R, Lecompte O, Poch O, Prieur D, Querellou J, Ripp R, Thierry JC, Van der OJ, Weissenbach J, Zivanovic Y, and Forterre P (2003). An integrated analysis of the genome of the hyperthermophilic archaeon *Pyrococcus abyssi*. *Mol Microbiol* 2003 Mar ;47 (6):1495 -512 47:1495-1512.
- Cole AR, Ji H, and Simpson RJ (2000). Proteomic analysis of colonic crypts from normal, multiple intestinal neoplasia and p53-null mice: a comparison with colonic polyps. *Electrophoresis* 21:1772-1781.
- Colombo S, D'Auria S, Fusi P, Zecca L, Raia CA, and Tortora P (1992). Purification and characterization of a thermostable carboxypeptidase from the extreme thermophilic archaeobacterium *Sulfolobus solfataricus*. *Eur J Biochem* 206:349-357.
- Contursi P, Cannio R, Prato S, Fiorentino G, Rossi M, and Bartolucci S (2003). Development of a genetic system for hyperthermophilic Archaea: expression of a moderate thermophilic bacterial alcohol dehydrogenase gene in *Sulfolobus solfataricus*. *FEMS Microbiol Lett* 2003 Jan 21;218 (1):115 -20 218:115-120.
- Corbett JM, Wheeler CH, and Dunn MJ (1995). Coelectrophoresis of cardiac tissue from human, dog, rat and mouse: towards the establishment of an integrated two-dimensional protein database. *Electrophoresis* 16:1524-1529.
- Curnow AW, Tumbula DL, Pelaschier JT, Min B, and Soll D (1998). Glutamyl-tRNA(Gln) amidotransferase in *Deinococcus radiodurans* may be confined to asparagine biosynthesis. *Proc Natl Acad Sci U S A* 95:12838-12843.
- De Vendittis E and Bocchini V (1996). Protein-encoding genes in the sulfothermophilic archaea *Sulfolobus* and *Pyrococcus*. *Gene* 176:27-33.

- Del Razo LM, Quintanilla-Vega B, Brambila-Colombres E, Calderon-Aranda ES, Manno M, and Albores A (2001). Stress proteins induced by arsenic. *Toxicol Appl Pharmacol* 177:132-148.
- DeLong EF and Pace NR (2001). Environmental diversity of bacteria and archaea. *Syst Biol* 50:470-478.
- DeRosa M, Gambacorta A, and Bu'lock JD (1975). Extremely thermophilic acidophilic bacteria convergent with *Sulfolobus acidocaldarius*. *J Gen Microbiol* 86:156-164.
- DeRosa M (1996). Archaeal lipids: structural features and supramolecular organization. *Thin Solid Films* 284-285:13-17.
- Ding YH, Zhang SP, Tomb JF, and Ferry JG (2002). Genomic and proteomic analyses reveal multiple homologs of genes encoding enzymes of the methanol:coenzyme M methyltransferase system that are differentially expressed in methanol- and acetate-grown *Methanosarcina thermophila*. *FEMS Microbiol Lett* 215:127-132.
- Dionne I, Robinson NP, McGeoch AT, Marsh VL, Reddish A, and Bell SD (2003). DNA replication in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Biochem Soc Trans* 2003 Jun ;31 (3):674 -6 31:674-676.
- Dunn MJ (1997). Quantitative two-dimensional gel electrophoresis: from proteins to proteomes. *Biochem Soc Trans* 25:248-254.
- Dunn MJ and Crisp SJ (1994). Detection of proteins in polyacrylamide gels using an ultrasensitive silver staining technique. *Methods Mol Biol* 32:113-8.:113-118.
- Easty D, Hart IR, Patel K, Seymour C, Yacoub M, Domscheit A, Gunther S, Postel W, Gorg A, and Dunn MJ (1991). Changes in protein expression during melanoma differentiation determined by computer analysis of 2-D gels. *Clin Exp Metastasis* 9:221-230.
- Eberhardt S, Korn S, Lottspeich F, and Bacher A (1997). Biosynthesis of riboflavin: an unusual riboflavin synthase of *Methanobacterium thermoautotrophicum*. *J Bacteriol* 179:2938-2943.
- Egelseer EM, Schocher I, Sleytr UB, and Sara M (1996). Evidence that an N-terminal S-layer protein fragment triggers the release of a cell-associated high-molecular-weight amylase in *Bacillus stearothermophilus* ATCC 12980. *J Bacteriol* 178:5602-5609.
- Fenn JB, Mann M, Meng CK, Wong SF, and Whitehouse CM (1989). Electrospray ionization for mass spectrometry of large biomolecules. *Science* 246:64-71.

- Fountoulakis M and Takacs B (2001). Effect of strong detergents and chaotropes on the detection of proteins in two-dimensional gels. *Electrophoresis* 22:1593-1602.
- Fusi P, Tedeschi G, Aliverti A, Ronchi S, Tortora P, and Gueritore A (1993). Ribonucleases from the extreme thermophilic archaebacterium *S. solfataricus*. *Eur J Biochem* 211:305-310.
- Gaasterland T, Maltsev N, Lobo J, and Chen GH (1994). Assigning function to CDS through qualified query answering: beyond alignment and motifs. *Proc Int Conf Intell Syst Mol Biol* 2:130-7.:130-137.
- Gamero-Castano M and Mora JF (2000). Kinetics of small ion evaporation from the charge and mass distribution of multiply charged clusters in electrosprays. *J Mass Spectrom* 35:790-803.
- Gebel TW (2001). Genotoxicity of arsenical compounds. *Int J Hyg Environ Health* 203:249-262.
- Gihring TM, Bond PL, Peters SC, and Banfield JF (2003). Arsenic resistance in the archaeon "*Ferroplasma acidarmanus*": new insights into the structure and evolution of the ars genes. *Extremophiles* 7:123-130.
- Giometti CS, Tollaksen SL, Mukund S, Zhou ZH, Ma K, Mai X, Adams MW (1995). *J. Chrom. A* 698: 341-349.
- Giometti CS, Reich C, Tollaksen S, Babnigg G, Lim H, Zhu W, Yates J, and Olsen G (2002). Global analysis of a "simple" proteome: *Methanococcus jannaschii*. *J Chromatogr B Analyt Technol Biomed Life Sci* 782:227-243.
- Gorg A (1999). IPG-Dalt of very alkaline proteins. *Methods Mol Biol* 112:197-209.
- Gorg A, Obermaier C, Boguth G, Harder A, Scheibe B, Wildgruber R, and Weiss W (2000). The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* 21:1037-1053.
- Gorg A, Postel W, and Gunther S (1988). The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* 9:531-546.
- Granoff A and Webster RG (1999). Viruses. In: *Encyclopedia of Virology*. New York: Academic Press 76-89.
- Grogan DW (1989). Phenotypic characterization of the archaebacterial genus *Sulfolobus*: comparison of five wild-type strains. *J Bacteriol* 171:6710-6719.

- Gulik A, Luzzati V, DeRosa M, and Gambacorta A (1988). Tetraether lipid components from a thermoacidophilic archaebacterium. Chemical structure and physical polymorphism. *J Mol Biol* 20:201:429-435.
- Gygi SP and Aebersold R (2000). Mass spectrometry and proteomics. *Curr Opin Chem Biol* 4:489-494.
- Gygi SP, Rochon Y, Franza BR, and Aebersold R (1999). Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol* 19:1720-1730.
- Hanford MJ and Peeples TL (2002). Archaeal tetraether lipids: unique structures and applications. *Appl Biochem Biotechnol* 97:45-62.
- Henry MP (1990). High-Performance Ion-Exchange Chromatography of Proteins. In: High Performance Liquid Chromatography in Biotechnology. New York: John Wiley and Sons
- Hershberger KL, Barns SM, Reysenbach AL, Dawson SC, and Pace NR (1996). Wide diversity of Crenarchaeota. *Nature* 384:420.
- Hillenkamp F, Karas M, Beavis RC, and Chait BT (1991). Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. *Anal Chem* 63:1193A-1203A.
- Hochstrasser DF (1998). Proteome in perspective. *Clin Chem Lab Med* 36:825-836.
- Hodges RS, Parker JM, Mant CT, and Sharma RR (1988). Computer simulation of high-performance liquid chromatographic separations of peptide and protein digests for development of size-exclusion, ion-exchange and reversed-phase chromatographic methods. *J Chromatogr* 458:147-67.:147-167.
- Hopmans EC, Schouten S, Pancost RD, van der Meer MT, and Sinninghe Damste JS (2000). Analysis of intact tetraether lipids in archaeal cell material and sediments by high performance liquid chromatography/atmospheric pressure chemical ionization mass spectrometry. *Rapid Commun Mass Spectrom* 14:585-589.
- Hough DW and Danson MJ (1999). Extremozymes. *Curr Opin Chem Biol* 3:39-46.
- Jaenicke R and Bohm G (1998). The stability of proteins in extreme environments. *Curr Opin Struct Biol* 8:738-748.
- James P (2000). Protein identification in the post-genome era: the rapid rise of proteomics. *Quarterly Review Biophys* 30:279-331.

- Jardine I (1990). Molecular weight analysis of proteins. *Methods Enzymol* 193:441-55.:441-455.
- Jonscher KR and Yates JR, III (1997). The quadrupole ion trap mass spectrometer--a small solution to a big challenge. *Anal Biochem* 244:1-15.
- Jonuscheit M, Martusewitsch E, Stedman KM, and Schléper C (2003). A reporter gene system for the hyperthermophilic archaeon *Sulfolobus solfataricus* based on a selectable and integrative shuttle vector. *Mol Microbiol* 2003 Jun ;48 (5):1241 -52 48:1241-1252.
- Josic D, Horn H, Schulz P, Schwinn H, and Britsch L (1998). Size-exclusion chromatography of plasma proteins with high molecular masses. *J Chromatogr A* 20:796:289-298.
- Kardinahl S, Schmidt CL, Hansen T, Anemuller S, Petersen A, Schafer G (1999). The strict molybdate-dependence of glucose-degradation by the thermoacidophile *Sulfolobus acidocaldarius* reveals the first crenarchaeotic molybdenum containing enzyme - an aldehyde oxidoreductase. *Eur. J. Biochem* 260: 540-548.
- Karger BL, Hancock WS, and ed (1996). *Methods in Enzymology*. San Diego: Academic Press, v177.
- Kawarabayasi Y, Hino Y, Horikawa H, Jin-no K, Takahashi M, Sekine M, Baba S, Ankai A, Kosugi H, Hosoyama A, Fukui S, Nagai Y, Nishijima K, Otsuka R, Nakazawa H, Takamiya M, Kato Y, Yoshizawa T, Tanaka T, Kudoh Y, Yamazaki J, Kushida N, Oguchi A, Aoki K, Masuda S, Yanagii M, Nishimura M, Yamagishi A, Oshima T, and Kikuchi H (2001). Complete genome sequence of an aerobic thermoacidophilic crenarchaeon, *Sulfolobus tokodaii* strain7. *DNA Res* 8:123-140.
- Kim J, Kim SH, Lee SU, Ha GH, Kang DG, Ha NY, Ahn JS, Cho HY, Kang SJ, Lee YJ, Hong SC, Ha WS, Bae JM, Lee CW, and Kim JW (2002). Proteome analysis of human liver tumor tissue by two-dimensional gel electrophoresis and matrixassisted laser desorption/ionization-mass spectrometry for identification of disease-related proteins. *Electrophoresis* 23:4142-4156.
- Koval SF and Hynes SH (1991). Effect of paracrystalline protein surface layers on predation by *Bdellovibrio bacteriovorus*. *J Bacteriol* 173:2244-2249.
- Kovtoun SV, English RD, and Cotter RJ (2002). Mass correlated acceleration in a reflectron MALDI TOF mass spectrometer: an approach for enhanced resolution over a broad mass range. *J Am Soc Mass Spectrom* 13:135-143.

- Krishna RG and Wold F (1993). Post-translational modification of proteins. *Adv Enzymol Relat Areas Mol Biol* 67:265-298.
- Kruft V and Wittmann-Liebold B (1991). Determination of peptide regions on the surface of the eubacterial and archaebacterial ribosome by limited proteolytic digestion. *Biochemistry* 30: 11781-11787.
- Kyrpides NC, Ouzounis CA, and Sander V. Novel protein families in Archaeal genomes. *Nucleic Acids Res.* 2003. In Press
- Ladenstein R and Antranikian G (1998). Proteins from hyperthermophiles: stability and enzymatic catalysis close to the boiling point of water. *Adv Biochem Eng Biotechnol* 61:37-85.:37-85.
- Laemmli UK (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227:680-685.
- Langner HW, Jackson CR, McDermott TR, and Inskeep WP (2001). Rapid oxidation of arsenite in a hot spring ecosystem, Yellowstone National Park. *Environ Sci Technol* 35:3302-3309.
- Laub MT, McAdams HH, Feldblyum T, Fraser CM, and Shapiro L (2000). Global analysis of the genetic network controlling a bacterial cell cycle. *Science* 290:2144-2148.
- Link AJ (1999). Autoradiography of 2-D gels. *Methods Mol Biol* 112:285-290.
- Loo JA, Brown J, Critchley G, Mitchell C, Andrews PC, and Ogorzalek Loo RR (1999a). High sensitivity mass spectrometric methods for obtaining intact molecular weights from gel-separated proteins. *Electrophoresis* 20:743-748.
- Loo JA, DeJohn DE, Du P, Stevenson TI, and Ogorzalek LR (1999b). Application of mass spectrometry for target identification and characterization. *Med Res Rev* 19:307-319.
- Lopez MF, Berggren K, Chernokalskaya E, Lazarev A, Robinson M, and Patton WF (2000). A comparison of silver stain and SYPRO Ruby Protein Gel Stain with respect to protein detection in two-dimensional gels and identification by peptide mass profiling. *Electrophoresis* 21:3673-3683.
- Lorite MJ, Tachil J, Sanjuan J, Meyer O, and Bedmar EJ (2000). Carbon monoxide dehydrogenase activity in *Bradyrhizobium japonicum*. *Appl Environ Microbiol* 2000 May ;66 (5):1871 -6 66:1871-1876.

- Luzzati V, Gambacorta A, DeRosa M, and Gulik A (1987). Polar lipids of thermophilic prokaryotic organisms: chemical and physical structure. *Annu Rev Biophys Chem* 16:25-47.:25-47.
- Mann M and Jensen ON (2003). Proteomic analysis of post-translational modifications. *Nat Biotechnol* 21:255-261.
- Mant CT and Hodges RS (2002). Reversed-phase liquid chromatography as a tool in the determination of the hydrophilicity/hydrophobicity of amino acid side-chains at a ligand-receptor interface in the presence of different aqueous environments. I. Effect of varying receptor hydrophobicity. *J Chromatogr A* 972:45-60.
- Mant CT, Parker JM, and Hodges RS (1987). Size-exclusion high-performance liquid chromatography of peptides. Requirement for peptide standards to monitor column performance and non-ideal behaviour. *J Chromatogr* 397:99-112.:99-112.
- Martusewitsch E, Sensen CW, and Schleper C (2000). High spontaneous mutation rate in the hyperthermophilic archaeon *Sulfolobus solfataricus* is mediated by transposable elements. *J Bacteriol* 182:2574-2581.
- Mayr J, Lupas A, Kellermann J, Eckerskorn C, Baumeister W, and Peters J (1996). A hyperthermostable protease of the subtilisin family bound to the surface layer of the archaeon *Staphylothermus marinus*. *Curr Biol* 6:739-749.
- Menendez C, Bauer Z, Huber H, Gadon N, Stetter KO, and Fuchs G (1999). Presence of acetyl coenzyme A (CoA) carboxylase and propionyl-CoA in autotrophic Crenarchaeota and indication of a 3-hydroxypropionate cycle in autotrophic carbon fixation. *J Bacteriol* 181:1088-1098.
- Meyer O, Frunzke K, Gadkari D, Jacobitz S, Hugendieck I, and Kraut M (1990). Utilization of carbon monoxide by aerobes: recent advances. *FEMS Microbiol Rev* 87:253-260.
- Meyer O, Gremer L, Ferner R, Ferner M, Dobbek H, Gnida M, Meyer-Klaucke W, and Huber R (2000). The role of Se, Mo and Fe in the structure and function of carbon monoxide dehydrogenase. *Biol Chem* 2000 Sep -Oct ;381 (9 -10):865 -76 381:865-876.
- Meyer O, Jacobitz S, and Kruger B (1986). Biochemistry and physiology of aerobic carbon monoxide-utilizing bacteria. *FEMS Microbiol Rev* 39:161-179.
- Minowa T, Ohtsuka S, Sasai H, and Kamada M (2000). Proteomic analysis of the small intestine and colon epithelia of adenomatous polyposis coli gene-mutant mice by two-dimensional gel electrophoresis. *Electrophoresis* 21:1782-1786.

- Molloy MP (2000). Two-dimensional electrophoresis of membrane proteins using immobilized pH gradients. *Anal Biochem* 280:1-10.
- Mortz E, Krogh TN, Vorum H, and Gorg A (2001). Improved silver staining protocols for high sensitivity protein identification using matrix-assisted laser desorption/ionization-time of flight analysis. *Proteomics* 1:1359-1363.
- Nadal M, Mirambeau G, Forterre P, Reiter W, and Duguet M (1986). Positively supercoiled DNA in a virus-like particle of an archaebacterium. *Nature* 321:256-258.
- Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Nelson WC, Ketchum KA, McDonald L, Utterback TR, Malek JA, Linher KD, Garrett MM, Stewart AM, Cotton MD, Pratt MS, Phillips CA, Richardson D, Heidelberg J, Sutton GG, Fleischmann RD, Eisen JA, Fraser CM, and . (1999). Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature* 399:323-329.
- Neter J, Kutner MH, Nachtsheim CJ, and Wasserman W (1996). *Applied Linear Statistical Models*. Boston: McGraw-Hill, 725-731.
- Neuhoff V, Arold N, Taube D, and Ehrhardt W (1988). Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis* 9:255-262.
- Neuhoff V, Stamm R, Pardowitz I, Arold N, Ehrhardt W, and Taube D (1990). Essential problems in quantification of proteins following colloidal staining with coomassie brilliant blue dyes in polyacrylamide gels, and their solution. *Electrophoresis* 11:101-117.
- Nguyen DN, Becker GW, and Riggin RM (1995). Protein mass spectrometry: applications to analytical biotechnology. *Journal of Chromatography A* 705:21-45.
- Nishihara JC and Champion KM (2002). Quantitative evaluation of proteins in one- and two-dimensional polyacrylamide gels using a fluorescent stain. *Electrophoresis* 23:2203-2215.
- Oberson J, Rawyler A, Brandle R, Canevascini G (1999) Analysis of the heat-shock response displayed by two *Chaetomium* species originating from different thermal environments. *Fungal. Genet. Biol.* 26: 178-189.
- O'Farrell PH (1975). High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* 250:4007-4021.

- Ogorzalek LR, Mitchell C, Stevenson TI, Martin SA, Hines WM, Juhasz P, Patterson DH, Peltier JM, Loo JA, and Andrews PC (1997). Sensitivity and mass accuracy for proteins analyzed directly from polyacrylamide gels: implications for proteome mapping. *Electrophoresis* 18:382-390.
- Pace NR (2001). The universal nature of biochemistry. *Proc Natl Acad Sci U S A* 98:805-808.
- Pappin DJ (1997). Peptide mass fingerprinting using MALDI-TOF mass spectrometry. *Methods Mol Biol* 64:165-73.:165-173.
- Pappin DJ (2003). Peptide mass fingerprinting using MALDI-TOF mass spectrometry. *Methods Mol Biol* 211:211-9.:211-219.
- Park ZY and Russell DH (2001). Identification of individual proteins in complex protein mixtures by high-resolution, high-mass-accuracy MALDI TOF-mass spectrometry analysis of in-solution thermal denaturation/enzymatic digestion. *Anal Chem* 73:2558-2564.
- Patton WF (1995). Biologist's perspective on analytical imaging systems as applied to protein gel electrophoresis. *J Chromatogr A* 55-87.
- Patton WF (2000). A thousand points of light: the application of fluorescence detection technologies to two-dimensional gel electrophoresis and proteomics. *Electrophoresis* 21:1123-1144.
- Patton WF (2002). Detection technologies in proteome analysis. *J Chromatogr B Analyt Technol Biomed Life Sci* 771:3-31.
- Pinheiro JC and Bates DM (2000). *Mixed Effects Models in S and S-PLUS*. New York: Springer.
- Prangishvili D, Stedman K, and Zillig W (2001). Viruses of the extremely thermophilic archaeon *Sulfolobus*. *Trends in Microbiology* 9:39-43.
- Prangishvili D, Holz I, Stieger E, Nickell S, Kristjansson JK, and Zillig W (2000). Sulfolobins, specific proteinaceous toxins produced by strains of the extremely thermophilic archaeal genus *Sulfolobus*. *J Bacteriol* 182:2985-2988.
- Quadroni M and James P (1999). Proteomics and automation. *Electrophoresis* 20:664-677.
- Rabilloud T (1990). Mechanisms of protein silver staining in polyacrylamide gels: a 10-year synthesis. *Electrophoresis* 11:785-794.

- Rabilloud T (1996). Solubilization of proteins for electrophoretic analyses. *Electrophoresis* 17:813-829.
- Rabilloud T (1998). Use of thiourea to increase the solubility of membrane proteins in two-dimensional electrophoresis. *Electrophoresis* 19:758-760.
- Rabilloud T (1999). Solubilization of proteins in 2-D electrophoresis. An outline. *Methods Mol Biol* 112:9-19.:9-19.
- Rabilloud T (2000). Detecting proteins separated by 2-D gel electrophoresis. *Anal Chem* 72:48A-55A.
- Rabilloud T (2002). Two-dimensional gel electrophoresis in proteomics: old fashioned, but it still climbs up the mountains. *Proteomics* 2:3-10.
- Rabilloud T, Adessi C, Giraudel A, and Lunardi J (1997). Improvement of the solubilization of proteins in two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* 18:307-316.
- Rabilloud T, Vuillard L, Gilly C, and Lawrence JJ (1994a). Silver-staining of proteins in polyacrylamide gels: a general overview. *Cell Mol Biol (Noisy -le-grand)* 40:57-75.
- Rabilloud T, Valette C, and Lawrence JJ (1994b). Sample application by in-gel rehydration improves the resolution of two-dimensional electrophoresis with immobilized pH gradients in the first dimension. *Electrophoresis* 15:1552-1558.
- Raman B, Cheung A, and Marten MR (2002). Quantitative comparison and evaluation of two commercially available, two-dimensional electrophoresis image analysis software packages, Z3 and Melanie. *Electrophoresis* 23:2194-2202.
- Reiter WD, Palm P, Henschen A, Lottspeich F, Zillig W, and Grampp B (1987). Identification and characterization of a defective SSV1 genome integrated into a tRNA gene in the archaeobacterium *Sulfolobus* sp. B12. *Mol Gen Genet* 206:144-153.
- Reysenbach AL, Ehringer M, and Hershberger K (2000). Microbial diversity at 83 degrees C in Calcite Springs, Yellowstone National Park: another environment where the Aquificales and "Korarchaeota" coexist. *Extremophiles* 4:61-67.
- Rice G, Stedman K, Snyder J, Wiedenheft B, Willits D, Brumfield S, McDermott T, and Young MJ (2001). Viruses from extreme thermal environments. *Proc Natl Acad Sci U S A* 98:13341-13345.

- Roepstorff P (2000). MALDI-TOF mass spectrometry in protein chemistry. *EXS* 88:81-97.
- Rose K, Kocher HP, Blumberg BM, and Kolakofsky D (1984). An improved procedure, involving mass spectrometry, for N-terminal amino acid sequence determination of proteins which are N alpha-blocked. *Biochem J* 217:253-257.
- Rothschild LJ and Mancinelli RL (2001). Life in extreme environments. *Nature* 409:1092-1101.
- Ruepp A, Graml W, Santos-Martinez ML, Koretke KK, Volker C, Mewes HW, Frishman D, Stocker S, Lupas AN, and Baumeister W (2000). The genome sequence of the thermoacidophilic scavenger *Thermoplasma acidophilum*. *Nature* 2000 Sep 28 ; 407 (6803):508 -13 407:508-513.
- Salmi J, Aittokallio T, Westerholm J, Griese M, Rosengren A, Nyman TA, Lahesmaa R, and Nevalainen O (2002). Hierarchical grid transformation for image warping in the analysis of two-dimensional electrophoresis gels. *Proteomics* 2:1504-1515.
- Sanchez JC, Hochstrasser D, and Rabilloud T (1999). In-gel sample rehydration of immobilized pH gradient. *Methods Mol Biol* 112:221-5.:221-225.
- Sara M and Sleytr UB (2000). S-Layer proteins. *J Bacteriol* 182:859-868.
- Schafer G (1996). Bioenergetics of the archaebacterium *Sulfolobus*. *Biochim Biophys Acta* 1277:163-200.
- Schafer G (1999). How can archaea cope with extreme acidity? *Novartis Found Symp* 221:131-44; discussion 145-51.:131-144.
- Schafer G, Purschke WG, Gleissner M, and Schmidt CL (1996). Respiratory chains of archaea and extremophiles. *Biochim Biophys Acta* 1275:16-20.
- Scheler C, Lamer S, Pan Z, Li XP, Salnikow J, and Jungblut P (1998). Peptide mass fingerprint sequence coverage from differently stained proteins on two-dimensional electrophoresis patterns by matrix assisted laser desorption/ionization-mass spectrometry (MALDI-MS). *Electrophoresis* 19:918-927.
- Schleper C, Kubo K, and Zillig W (1992). The particle SSV1 from the extremely thermophilic archaean *Sulfolobus* is a virus: demonstration of infectivity and of transfection with viral DNA. *Proc Natl Acad Sci U S A* 89:7645-7649.

- Schleper C, Roder R, Singer T, and Zillig W (1994). An insertion element of the extremely thermophilic archaeon *Sulfolobus solfataricus* transposes into the endogenous beta-galactosidase gene. *Mol Gen Genet* 243:91-96.
- Schubel U, Kraut M, Morsdorf G, and Meyer O (1995). Molecular characterization of the gene cluster *coxMSL* encoding the molybdenum-containing carbon monoxide dehydrogenase of *Oligotropha carboxidovorans*. *J Bacteriol* 177:2197-2203.
- Schuster B and Sleytr UB (2000). S-layer-supported lipid membranes. *J Biotechnol* 74:233-254.
- Schwartz H, Langin T, Platsch H, Richert J, Bomm S, Schmidt M, Hillen H, Blaschke G, Meyer J, Darius H, and Buerke M (2002). Two-dimensional analysis of myocardial protein expression following myocardial ischemia and reperfusion in rabbits. *Proteomics* 2:988-995.
- Seegerer AH, Burggraf S, Fiala G, Huber G, Huber R, Pley U, and Stetter KO (1993). Life in hot springs and hydrothermal vents. *Orig Life Evol Biosph* 23:77-90.
- Sehgal AC, Tompson R, Cavanagh J, and Kelly RM (2002). Structural and catalytic response to temperature and cosolvents of carboxylesterase EST1 from the extremely thermoacidophilic archaeon *Sulfolobus solfataricus* P1. *Biotechnol Bioeng* 80:784-793.
- Sensen CW, Charlebois RL, Chow C, Clausen IG, Curtis B, Doolittle WF, Duguet M, Erauso G, Gaasterland T, Garrett RA, Gordon P, de Jong IH, Jeffries AC, Kozera C, Medina N, De Moors A, Van der OJ, Phan H, Ragan MA, Schenk ME, She Q, Singh RK, and Tolstrup N (1998). Completing the sequence of the *Sulfolobus solfataricus* P2 genome. *Extremophiles* 2:305-312.
- She Q, Singh RK, Confalonieri F, Zivanovic Y, Allard G, Awayez MJ, Chan-Weiher CC, Clausen IG, Curtis BA, De Moors A, Erauso G, Fletcher C, Gordon PM, Heikamp-de Jong I, Jeffries AC, Kozera CJ, Medina N, Peng X, Thi-Ngoc HP, Redder P, Schenk ME, Theriault C, Tolstrup N, Charlebois RL, Doolittle WF, Duguet M, Gaasterland T, Garrett RA, Ragan MA, Sensen CW, and Van der OJ (2001). The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc Natl Acad Sci U S A* 98:7835-7840.
- Sickmann A, Marcus K, Schafer H, Butt-Dorje E, Lehr S, Herkner A, Suer S, Bahr I, and Meyer HE (2001). Identification of post-translationally modified proteins in proteome studies. *Electrophoresis* 22:1669-1676.
- Silver S (1996). Bacterial resistances to toxic metal ions--a review. *Gene* 179:9-19.

- Soskic V and Godovac-Zimmermann J (2001). Improvement of an in-gel tryptic digestion method for matrix-assisted laser desorption/ionization-time of flight mass spectrometry peptide mapping by use of volatile solubilizing agents. *Proteomics* 1:1364-1367.
- Stedman KM, Schleper C, Rumpf E, and Zillig W (1999). Genetic requirements for the function of the archaeal virus SSV1 in *Sulfolobus solfataricus*: construction and testing of viral shuttle vectors. *Genetics* 152:1397-1405.
- Stensballe A and Jensen ON (2001). Simplified sample preparation method for protein identification by matrix-assisted laser desorption/ionization mass spectrometry: in-gel digestion on the probe surface. *Proteomics* 1:955-966.
- Switzer RC, III, Merrill CR, and Shifrin S (1979). A highly sensitive silver stain for detecting proteins and peptides in polyacrylamide gels. *Anal Biochem* 98:231-237.
- Tang TH, Rozhdestvensky TS, d'Orval BC, Bortolin ML, Huber H, Charpentier B, Branlant C, Bachellerie JP, Brosius J, and Huttenhofer A (2002). RNomics in Archaea reveals a further link between splicing of archaeal introns and rRNA processing. *Nucleic Acids Res* 30:921-930.
- Tekaia F, Yeramian E, and Dujon B (2002). Amino acid composition of genomes, lifestyles of organisms, and evolutionary trends: a global picture with correspondence analysis. *Gene* 297:51-60.
- Tolstrup N, Sensen CW, Garrett RA, and Clausen IG (2000). Two different and highly organized mechanisms of translation initiation in the archaeon *Sulfolobus solfataricus*. *Extremophiles* 4:175-179.
- Trent JD, Gabrielsen M, Jensen B, Neuhaard J, and Olsen J (1994). Acquired thermotolerance and heat shock proteins in thermophiles from the three phylogenetic domains. *J Bacteriol* 176:6148-6152.
- van de Vossenberg JL, Driessen AJ, Grant WD, and Konings WN (1999). Lipid membranes from halophilic and alkali-halophilic Archaea have a low H⁺ and Na⁺ permeability at high salt concentration. *Extremophiles* 3:253-257.
- van de Vossenberg JL, Driessen AJ, and Konings WN (1998b). The essence of being extremophilic: the role of the unique archaeal membrane lipids. *Extremophiles* 2:163-170.
- van de Vossenberg JL, Ubbink-Kok T, Elferink MG, Driessen AJ, and Konings WN (1995). Ion permeability of the cytoplasmic membrane limits the maximum growth temperature of bacteria and archaea. *Mol Microbiol* 18:925-932.

- Voss T and Haberl P (2000). Observations on the reproducibility and matching efficiency of two-dimensional electrophoresis gels: consequences for comprehensive data analysis. *Electrophoresis* 21:3345-3350.
- Vuillard L, Braun-Breton C, and Rabilloud T (1995). Non-detergent sulphobetaines: a new class of mild solubilization agents for protein purification. *Biochem J* 305:337-343.
- Vuillard L, Rabilloud T, and Goldberg ME (1998). Interactions of non-detergent sulfobetaines with early folding intermediates facilitate in vitro protein renaturation. *Eur J Biochem* 256:128-135.
- Wasinger VC, Cordwell SJ, Cerpa-Poljak A, Yan JX, Gooley AA, Wilkins MR, Duncan MW, Harris R, Williams KL, and Humphery-Smith I (1995). Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*. *Electrophoresis* 16:1090-1094.
- Watanabe Y, Yokobori S, Inaba T, Yamagishi A, Oshima T, Kawarabayasi Y, Kikuchi H, and Kita K (2002). Introns in protein-coding genes in Archaea. *FEBS Lett* 510:27-30.
- White D (2000). The Physiology and Biochemistry of Prokaryotes . New York: Oxford University Press.
- Wilce MC, Aguilar MI, and Hearn MT (1993). High-performance liquid chromatography of amino acids, peptides and proteins. CXXII. Application of experimentally derived retention coefficients to the prediction of peptide retention times: studies with myohemerythrin. *J Chromatogr* 19;632:11-18.
- Wirth HJ, Unger KK, and Hearn MT (1991). High-performance liquid chromatography of amino acids, peptides and proteins. CIX. Investigations on the relation between the ligand density of cibacron blue immobilized porous and non-porous sorbents and protein-binding capacities and association constants. *J Chromatogr* 550:383-395.
- Yan Z, Fujiwara S, Kohda K, Takagi M, and Imanaka T (1997). In vitro stabilization and in vivo solubilization of foreign proteins by the beta subunit of a chaperonin from the hyperthermophilic archaeon *Pyrococcus* sp. strain KOD1. *Appl Environ Microbiol* 63:785-789.
- Yates JR, III (1998). Mass spectrometry and the age of the proteome. *J Mass Spectrom* 33:1-19.

- Zillig W, Stetter KO, Wunderl S, Schulz W, Priess H, and Scholz I (1980). The *Sulfolobus* "Caldariella" Group: Taxonomy on the basis of the structure of DNA-dependent RNA polymerases. *Arch Microbiol* 125:259-269.
- Zillig W, Arnold HP, Holz I, Prangishvili D, Schweier A, Stedman K, She Q, Phan H, Garrett R, and Kristjansson JK (1998). Genetic elements in the extremely thermophilic archaeon *Sulfolobus*. *Extremophiles* 2:131-140.
- Zuo X and Speicher DW (2000). Quantitative evaluation of protein recoveries in two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* 21:3035-3047.

MONTANA STATE UNIVERSITY - BOZEMAN



3 1762 10391960 9