GENOME-WIDE ANALYSIS OF MULTIPROTEIN COMPLEXES AND METALLOPROTEINS IN THE HYPERTHERMOPHILIC ARCHAEON *PYROCOCCUS FURIOSUS*

by

JOSEPH WALKER SCOTT

(Under the Direction of Michael W. W. Adams)

ABSTRACT

Multiprotein complexes and metalloproteins largely determine the metabolic potential of an organism. The genes that encode multiprotein complexes and metalloproteins in any organism are largely unknown. Predicting what these genes are from genome sequence data is virtually impossible due to great diversity of amino acids sequences that are involved in protein oligomerization and metal binding. In this dissertation, an experimental method is presented that circumvents this problem by the direct purification of multiprotein complexes and metalloproteins from native microbial biomass. The biomass of a model microorganism, *Pyrococcus furiosus,* was fractionated by non-denaturing multistep chromatography and chromatographic fractions were analysed by mass spectrometry to identify proteins that co-eluted in common fractions. Potential multiprotein complex identification was based on the cofractionation of proteins encoded by adjacent genes in the *P. furiosus* genome. A total of 106 potential heteromeric multiprotein complexes were identified, of which only 20 were already known or predicted. One of them, PC-81 (PF1838/PF1837) was of interest because of its homology to two previously purified heterotetrameric ($\alpha_2\beta_2$) enzymes termed ACS1

(PF1540/PF1787) and ACS2 (PF1540/PF1837). Since PC-81 and ACS2 had a common β-subunit (PF1837), we recombinantly expressed in *E. coli* all ten ACS complexes from the five α subunits and two β subunits encoded in the *P. furiosus* genome. All ten enzymes were active with varying but overlapping substrate specificities. Metalloproteins were identified in a similar manner except that fractions were analyzed by ICP-MS in addition to peptide mass spectrometry. Of the 343 metal peaks identified during biomass fractionation, 158 were not associated with a known metalloprotein. Two metalloproteins were purified to homogeneity. One of them, PF0086 or alanyl-tRNA synthetase COOH terminus had been predicted to contain zinc. However, it co-purified with nickel rather than a zinc peak through three chromatography steps. The other, PF1972 or anaerobic ribonucleotide reductase activase, is a known iron protein that very surprisingly co-purified with molybdenum as well as iron through five chromatography steps. This method therefore identified one new nickel and one new molybdenum-containing protein from *P. furiosus*, in addition to approximately 100 new multiprotein complexes, thereby demonstrating the general applicability of this approach.

INDEX WORDS: Hyperthermophilic archaea, *Pyrococcus furiosus*, multiprotein complexes, metalloproteome, nickel, molybdenum, proteomics, genomics, acetyl-CoA synthetase, alanyl-tRNA synthetase, anaerobic ribonucleotide reductase activase, glycolysis, gluconeogenesis, GTP-dependent phophoenolpyurvate carboxykinase, ATP, GTP, biomass fractionation, mass spectrometry, ICP-MS

GENOME-WIDE ANALYSIS OF MULTIPROTEIN COMPLEXES AND METALLOPROTEINS IN THE HYPERTHERMOPHILIC ARCHAEON *PYROCOCCUS FURIOSUS*

by

JOSEPH WALKER SCOTT

B.S., University of Arkansas, 1998

M.S., University of Florida, 2002

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2014

GENOME-WIDE ANALYSIS OF MULTIPROTEIN COMPLEXES AND METALLOPROTEINS IN THE HYPERTHERMOPHILIC ARCHAEON *PYROCOCCUS FURIOSUS*

by

JOSEPH WALKER SCOTT

Major Professor:	Michael W. W. Adams
Committee:	Michael K. Johnson Robert. A. Scott William B. Whitman

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2014

DEDICATION

This dissertation is dedicated to Janie Walker.

ACKNOWLEDGEMENTS

I must acknowledge those who encouraged me to continue when I did not think I even cared anymore. Names that readily come to mind are Dr. Michael W. W. Adams, Dr. Angeli Lal Menon, and Mr. Farris Poole II.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Genomics.

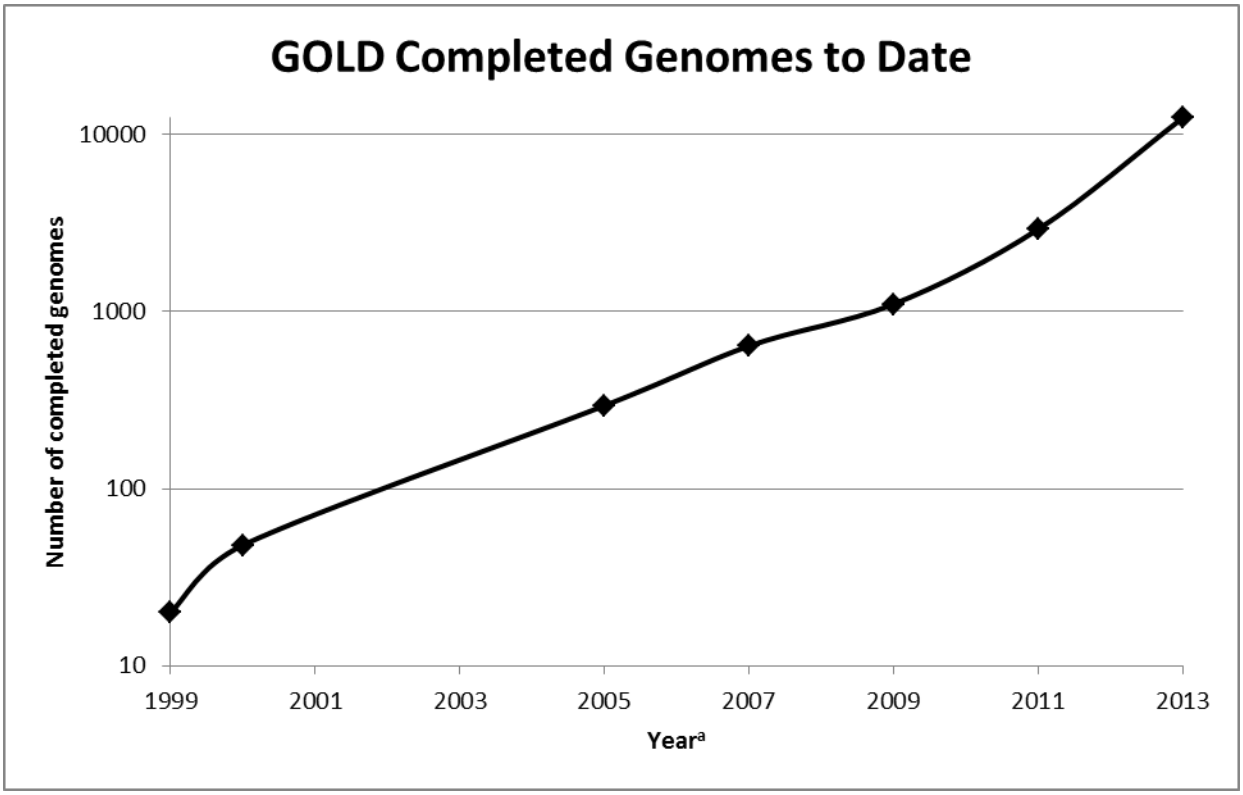
The amount of resources in money, people, and time invested into sequencing projects over the years reflects an expectation of a tremendous return. The number of sequenced genomes increases exponentially each year (Figure 1.1). Over 12,000 genomes of individual organisms have been sequenced with numerous, additional, sequencing projects underway (Genomes Online Database). Sequencing of the human genome was predicted to take 15 years and to cost $3 billion [1]. It involved an international effort from 20 research institutions in the United States, United Kingdom, France, Germany, Japan, and China. The expected return from sequencing the human genome was knowledge of the chromosomal location of every human gene, the nucleotide sequence of those genes, and the relationship of the genes to disease. Ideally, this knowledge would be helpful in assessing an individual's susceptibility to diseases while also predicting a patient's response to medical treatments. A question more fundamental than those concerning the relationship between human health and genetic make-up is the underlying motivation that propels not only genomic studies, but all studies in the life sciences. That question is: How does life maintain itself? The genome has been called "the blueprint of life" [2]. This lofty designation suggests that DNA contains the information to build all the parts necessary for the maintenance of life. Therefore, in theory a study of an organism's genome alone would facilitate prediction of the life processes in a biological system based upon the principle of genes. However, since the completion of the map of the human genome in 2003, some have suggested that many statements concerning the promise of genomics have been

Figure 1.1

Completed genomes in Genome On-line Database by year.

[a]References for given years. 1999 [3]; 2000 [4]; 2005 [5]; 2007 [6]; 2009 [7]; 2011 [8]

2013 Number taken from Genomes On-line Database on 12/27/2013.


GOLD Completed Genomes to Date
Number of completed genomes
10000
1000
100
10
1999
2001
2003
2005
2007
2009
2011
2013
Year[a]

premature. Eric Lander of the Whitehead institute has said: *We've called the genome the blueprint, the holy grail, all sorts of things. But it's only a parts list. If I gave you the parts list for the Boeing 777 and it has 100,000 parts, I don't think you could screw it together, and you certainly wouldn't understand why it flew* [9]. There are several reasons why the promised benefits of genomics have been delayed. Some of which are 1) the functions of roughly one-half of the genes in *E. coli* and *S. cerevisiae* (arguably the best studied organisms whose genomes have been sequenced) have not been experimentally verified making it difficult to draw accurate conclusions concerning the complex nature of biological systems [10-12]. 2) As of yet it is difficult to discern sequence information concerning interaction between individual gene products that form functional units. 3) Related to the previous point is the difficulty in discerning post translational modifications such as phosphorylation, glycosylation, acetylation, cleavage etc. that affects function. 4) Life is more than DNA. Life exists as an interplay between DNA, its gene products, and the environment. Gene products may require certain temperatures, proton concentrations, or small inorganic molecules such as salts or metal ions to carry out biological functions which are needed to sustain life. The increasing number of sequenced genomes has prompted a trend toward using genomic structure as a basis in formulating and answering questions about cellular processes. This shift is a defining feature of what has been termed the post-genomic era [13].

An obvious advantage to having the sequences of whole genomes available is that it precludes the necessity to purify a protein and determine its amino acid sequence before cloning the gene that expresses it. A simple search of a sequence database allows one to bypass those steps and begin a cloning project by designing oligonuclotide primers based on the DNA sequence of the gene of interest. Genome sequencing has also allowed for the development of

sub-disciplines with the common suffix of –omics which denotes their relationship to genomics. Two common examples of -omics studies are transcriptomics and proteomics. Transcriptomics is a discipline which involves the analysis of the RNAs expressed by a specific cell type or organism at specific times or under a specific set of conditions. Likewise, proteomics studies the complete set or subset of proteins that are expressed temporally and spatially. The –omics suffix can be appended to any root that describes a collection of any biologically relevant molecules (Table 1.1). Each of these disciplines complements genomics by helping to tease out higher biological meaning from sequence data. A chief characteristic of -omics technologies is the generation of a lot of data in a relatively short amount of time. For example whole-genome transcriptomics experiments in the form of microarrays can be used to analyze thousands of genes spotted onto chips which are used to disclose the expression patterns of all genes in a particular genome. Likewise, in proteomics experiments, mass spectrometry can detect a multitude of proteins expressed in a tissue, cell type, organelle, bacterium etc. Again, these proteomics experiments are dependent on the completed genomic sequences of the organism under study since the peptide mass fingerprint is searched against a data base of virtually digested proteins that have been deduced from the genome sequence of the host organisms. The high-throughput nature of the so called '-omics revolution' has produced a study of biology that increasingly has the salient feature of being data rather than hypothesis driven. Some in the scientific community have expressed sentiments that data driven science has no merit [14]. Nobel Prize winning scientist, Arthur Kornberg does not go as far as suggesting genomics is useless, however he bemoans what he sees as biochemistry 'becoming less fashionable' in the post genomic era [15]. He states that the purification and characterization of enzymes from a cell-free system "does something that proteomic and other 'omics' cannot yet do." Although

Table 1.1 Example of –omics technologies

Discipline	Subject of study
Genomics-	DNA sequences, Gene organization
Epigenomics-	DNA modifications, Modification of DNA binding proteins
Transcriptomics-	Expression profiles of coding RNA. Also includes the study of non-coding RNA
Proteomics-	Entire complement of proteins produced by an organism
Glycomics-	Entire complement of sugar chains produced by an organism
Lipidomics-	Entire complement of cellular lipids
Metabolomics-	Entire complement of metabolites produced by an organism.
Metalloproteomics-	Entire complement of metalloproteins produced by an organism

transcriptomic and proteomic experiments provide clues as to the functions of genes, the data generated is similar to genome sequence data in that there is plenty of it and it does not provide the information needed to discern the specific functions of the genes expressed.

On the other hand, it could be argued that a focus on the characterization of individual proteins is too narrow to be able to represent how that protein functions in a network thereby giving an incomplete picture of the role the protein plays in a cell [16]. The studies which are the subject of this work reflect a view that old-fashioned biochemistry and the methods used in the era of post genomics can be used in concert to achieve a comprehensive view of the relationship between the genome, its gene products and the environment in which they function.

Pyrococcus furiosus as a model organism.

The haploid human genome is composed of 23 chromosomes, three billion pairs of nucleotides and 20,000-30,000 protein encoding genes. Many of the genes are transcribed into mRNA that is subject to splicing events and produce proteins which are different than would be predicted from the DNA sequence. In addition, there are 210 distinct human cell types all having the same DNA (with a few exceptions), but expressing genes differentially. Furthermore, 98 % of the Human genome sequence is repeat elements, transposons and non-coding DNA and has been referred to as 'junk-DNA'. However, there have been developments that suggest possible functions for non-coding DNA [17]. It is clear from the complexity and mere size of the human genome that an effort to decipher it presents an almost insurmountable challenge for those researchers who dare to study it.

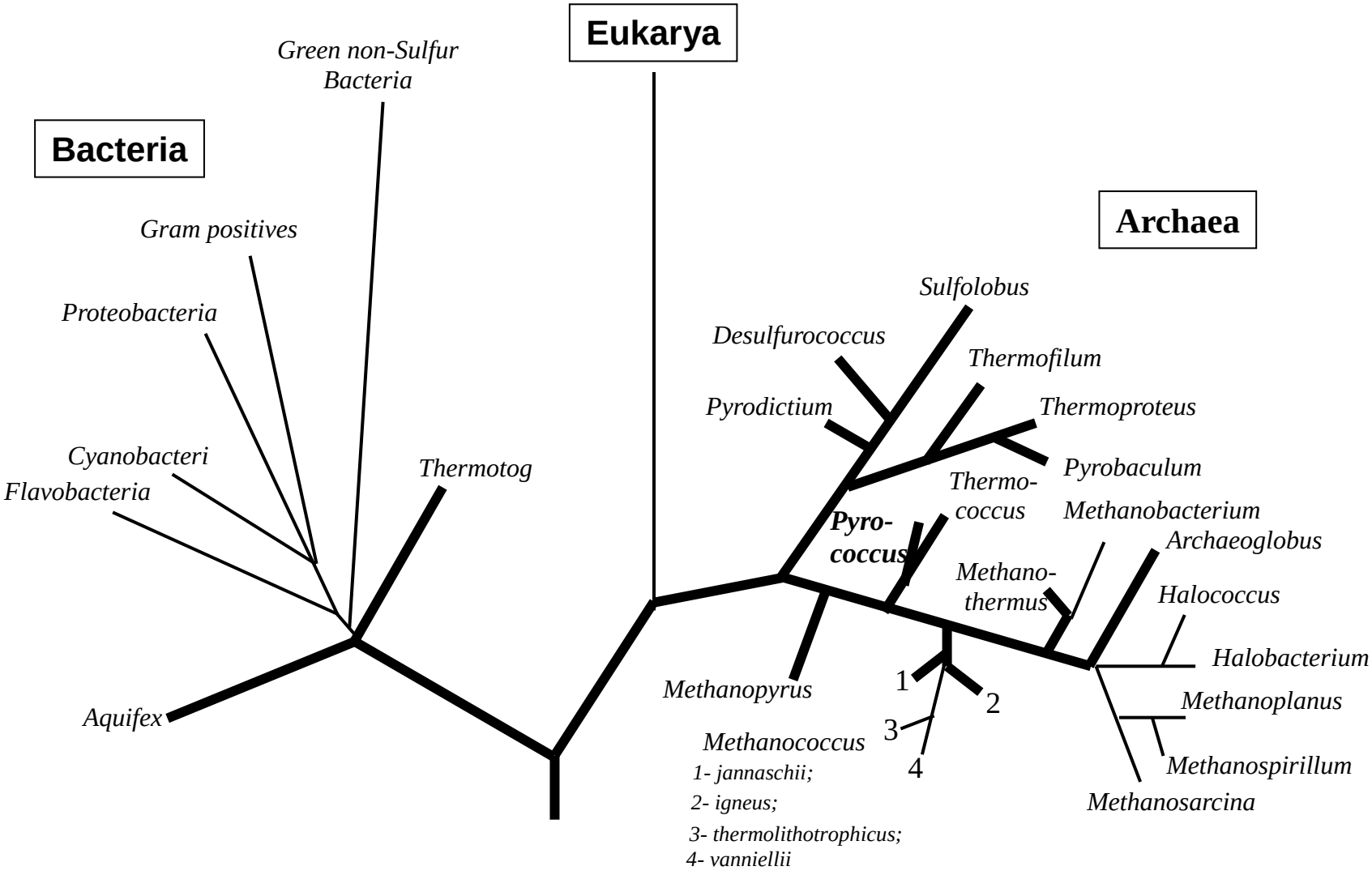
Pyrococcus furious, a much simpler species than *Homo sapiens*, is a single celled organism from the archaea domain of life. Interestingly the universal tree of life indicates that the archaea and eucarya domains share a common line of evolution from the last universal

common ancestor (LUCA), while the bacteria domain follows its own line (Figure 1.2). Thus although members of the bacteria and archaea domains are single celled prokaryotes, archaea are more evolutionarily related to the eucarya than to bacteria. *Pyrococcus furiosus* was originally isolated in 1986 from a shallow marine vent, but is commonly found in deep-sea vents [19]. *P. furiosus* is a hyperthermophile with an optimal growth temperature of 100°C and is known as the '*E. coli* of life at 100°C' because it is by far the best studied hyperthermophile. It has been the subject of studies regarding metabolism, gene organization, genetics, physiology, and enzyme characterization [20-24]. In fact, the experiments presented in this paper were performed in a laboratory that has used *P. furiosus* as a model organism for over 20 years. In this lab, *P. furiosus* has previously been the subject of a genomics project that aimed to determine the structures of all the proteins expressed in its genome [25-27]. The premise of this structural genomics project was based on the belief that proteins would provide structural information that could be correlated to function so that rather than depending solely on sequence to decipher the genome, protein structure could be used as well. One of the major hindrances in this effort was that only about 25 % of the open-reading-frames in a given prokaryotic genome will yield stable, soluble protein when expressed alone. This percentage does not include membrane proteins. One explanation for this attrition rate is that these proteins are subunits that need their partners for stability or that lack a post translational modification that is not performed in *E. coli*. The genome of *P. furiosus* was sequenced in 2001 as part of a DOE funded project and was found to have a size of 1.9 million base pairs which make up around 2100 open reading frames (ORF) [28]. 59 % or 1233 of those ORFs are members of over 400 basic transcriptional units called operons [29], a claim supported by ORFs either being within or overlapping by 16 nucleotides on the same strand [30; 31] and by their co-regulation under differing growth conditions as

Figure 1.2

Universal Phylogenetic Tree

The bold lines represent the hyperthermophilic genera. This phylogenetic tree is adapted from [18].


Eukarya
Green non-Sulfur Bacteria
Bacteria
Archaea
Gram positives
Proteobacteria
Cyanobacteri
Flavobacteria
Thermotog
Aquifex
Sulfolobus
Desulfurococcus
Thermofilum
Pyrodictium
Thermoproteus
Pyrobaculum
Thermo-coccus
Pyro-coccus
Methanobacterium
Archaeoglobus
Methano-thermus
Halococcus
Halobacterium
Methanoplanus
Methanospirillum
Methanosarcina
Methanopyrus
Methanococcus
1
2
3
4
1- jannaschii;
2- igneus;
3- thermolithotrophicus;
4- vanniellii

demonstrated from whole-genome transcriptomic analysis. In *P. furiosus*, most of the characterized protein complexes are encoded by adjacent genes. Thus, the small, well-defined, highly regulated genome of *P. furiosus* will aid in predicting the protein partners that are required to be co-expressed to produce stable soluble protein. However, a well-defined genome is not the same thing as a perfectly defined genome. That being said, there are some errors in the *P. furiosus* sequence that add complications. The original genome of *P. furiosus* was independently annotated by The Institute for Genomic Research (TIGR) and the National Center for Biotechnology Information (NCBI). More than 500 of the annotated genes differed significantly in size in the two databases [22]. Therefore, more direct experimental confirmation on a broad scale would be useful not only in defining ORFs but also in defining the operons that encode the proteins which make up multi-subunit complexes and modified proteins. Purification of proteins from native biomass is the most direct method to determine which protein complexes and protein modifications occur in the cell. *P. furiosus* can be routinely grown in kilogram quantities in readily accessible fermentors, a capability which provides a virtually limitless amount of biomass for protein purification and characterization [32].

The ability of *P. furiosus* to convert peptides and complex carbohydrates into hydrogen makes it attractive from a biomass conversion standpoint and its ability to live at boiling temperatures lends itself to the understanding of cellular processes at high temperatures. In addition, some believe that hyperthermophiles, such as *P. furious* are the most slowly evolving of all organisms and thus are representative of the most ancient organisms. It may be the case that the first life forms originated from deep sea vent environments which are at temperature at which CO_2 fixation takes place abiotically [33-35]. Thus a greater understanding of the *P. furiosus* genome could lend itself to a greater understanding of evolution.

Multiprotein complexes.

As our knowledge of protein structure increases, it is becoming more apparent that the majority of proteins function in complexes with other proteins, while only a small fraction function in isolation [36]. A lack of a protein partner during overexpression is a convincing argument as to why the Structural Genomics effort had a high attrition rate. It has been suggested that when a protein is in a complex, the possibility of denaturation is minimized due to a reduced surface area exposed to solvent [37].

On the other hand, multiple subunits also make it possible for complexes to replace parts of the complex if any subunit is damaged for a reason. If the enzyme was rather composed of a single polypeptide any damage might destroy the function of the whole protein. Multi-subunit complexes can form large molecular assemblies that function in cellular processes such as protein synthesis [38; 39], cell motility [40], and extracelluar degradation of plant biomass [41]. Some even larger or supramolecular assemblies function in what are called metabolons. These are temporary complexes of enzymes held together by noncovalent interactions and are involved in a common pathway which allows for the efficient channeling of substrates from the beginning to the end of the pathway [42-44]. Some molecular complexes are highly stable such as ribosomes, while others are less so and possibly participate in transient interactions that are important for signaling or regulatory networks. With this in mind, it becomes even more apparent why a greater knowledge of the nature of protein complexes in *P. furiosus* would be very helpful in learning more about the biology of this organism.

Protein complexes have been categorized in our lab into three types based upon the organization of the genes that express their component subunits [45]. The complexes may be either transient or permanent.

Type 1 complexes. These are heteromeric protein complexes encoded by at least two ORFs that are adjacent in the genome. The 2-keto acid oxidoreductases (KOR) of *P. furious* are examples of this type of complex [46-49]. These enzymes are involved in the oxidative decarboxylation of 2-keto acids which results in a coenzyme A derivative. The genome sequence of *P. furiosus* indicates that the KORs are organized into operons. Four have been purified and characterized [49]. Pyruvate ferredoxin oxidoreductase (POR), 2-ketoisovalerate ferredoxin oxidoreductase (VOR), and 2-ketoglutarate ferredoxin oxidoreductase (KGOR) have $\alpha_2\beta_2\gamma_2\delta_2$ quaternary structures. The γ subunit of POR and VOR is shared by those two complexes. Indole pyruvate ferredoxin oxidoreductase (IOR) has an $\alpha_2\beta_2$ quaternary structure. However, the subunits appear to be a fused mosaic of the homologous subunits of the other KORs. The *P. furiosus* genome also encodes a fifth KOR homolog with genes immediately downstream of KGOR. Microarray data show that this putative 2-keto acid ferredoxin oxidoreductase (XOR) is co-regulated with KGOR and shares its δ subunit with it much like POR shares its γ subunit with VOR. [50] An additional putative KOR homolog is encoded by the adjacent ORFs PF0753 and PF0754 [45]. This KOR has been partially purified from native *P. furiosus* biomass [45], however additional experimental data about this complex is unavailable. With only two genes encoding the complex, it may have a quaternary structure similar to IOR. It is evident that the genome sequence can be a useful aid in predicting protein complexes when working with an organism such as *P. furiosus* which has a small, well-defined, and highly regulated genome.

Type 2 complexes. The second type of complex that we have categorized is composed of subunits that are encoded by genes which are not adjacent in the genome and therefore not co-regulated in an operon [45]. Therefore, the genome sequence is not helpful in predicting these types of complexes. Examples of this type of complex are acetyl-CoA synthetases (ACS) 1 and

2 [51]. ACS1 catalyzes the last step in sugar fermentation and fermentation of some amino acids. ACS2 also catalyzes the last step in fermentation of some amino acids. These enzymes were purified by following enzyme activity. They were found to be composed of two different types of subunits in an $\alpha_2\beta_2$ quaternary organization. N-terminal sequencing was used to determine the amino acid sequences of the enzyme subunits. All subunits that made up the two enzymes had different amino acid sequences. After the genome of *P. furiosus* was sequenced, it was determined that ACS1 was composed of subunits encoded by PF1540 and PF1787 and ACS2 was made of subunits encoded by PF0532 and PF1837. An observation of the gene numbers makes it apparent that none of genes of either complex are adjacent in the genome.

Type 3 complexes. The third type of complex is composed of identical subunits. As opposed to heteromers, type 3 complexes are homomers which are composed of subunits encoded by a single gene. Given a polypeptide sequence, it is not clear what oligomerization state will be formed if any. There are several techniques that are frequently used which can help determine the oligomeric state. These include: analytical gel filtration chromatography, mass spectrometry, analytical ultracentrifugation, and dynamic light scattering among others [52].

Homooligomerization allows cells to make larger protein complexes without requiring an increase in genome size [36]. Most homooligomers are dimers although much larger complexes do exist. An intriguing example of a very large complex produced in *P. furiosus* is the sulfur induced protein A (Sip A) [53; 54]. PF2025 or *sipA* is upregulated by sulfide and elemental sulfur. It encodes a 19 kDa protein that aggregates into a homooligomer of up to approximately 100 MDa and is composed of about 5,000 identical subunits and up to 45,000 iron sulfur clusters. It has a unique, repetitive C-terminal tail that is required for polymerization [55].

However, the sequence in of itself provided no insight into its function. Of all three types of complexes described, genome structure has the most bearing in predicting type 1 complexes.

Metalloproteins.

A study of the nutritional requirements of every life form would suggest that metal ions are required for metabolic function. Bulk metals such as sodium, potassium, calcium, and magnesium are present in organisms in relatively large amounts [56]. Bulk metals can be used to maintain osmotic balance, carry charges across membranes, among other processes. The bulk metals are not necessarily bound to organic molecules when performing their functions although they often are. On the other hand, iron, zinc, copper, cobalt, nickel and other metals are present at trace levels in cells, tissues and biological fluid [57]. The trace amounts of these metals infer a function with proteins that are also present in 'trace' amounts [58]. The catalytic nature of enzyme proteins allows them to function at low concentrations. Proteins that incorporate metal ions into their structures are called metalloproteins. The assimilation and incorporation of metals into proteins is the basis of many life functions including electron transfer, catalysis, and protein structure stabilization. Some estimates suggest metalloproteins comprise up to 40 % of all proteins [59]. The designation of metalloprotein has typically been conferred to a protein after purification [60-62]. There are several informatics approaches to identify metalloproteins. However, these methods are limited in that they depend heavily on homology to characterized metalloproteins [60; 62-65]. One-third or more of genes in a typical cell encode proteins that have not been experimentally verified to exist [66] so it is not known whether or not these genes encode metalloproteins. The metal binding sites of most metalloproteins contribute only a small number of ligands to coordinate the metal [67]. Because of the three dimensional folding of proteins, placement of the metal amino acid ligands in the primary sequence can be varied to a

great degree. Therefore, it is very difficult to predict with available informatic methods which of the hypothetical proteins are metalloproteins.

Another impediment to more robust predictive capabilities may be biased sampling of the metalloproteins that have been structurally characterized [67]. During crystallization, crystals are often soaked in heavy metals to aid in solving the protein structure. Therefore, those proteins with binding sites that readily displace and exchange metals have an increased probability of being solved than those proteins that do not release their metal. Those binding sites that do readily exchange their metals may represent only a small proportion of the possible folds that do bind metals.

Because of the abundance of protein needed for crystallization, recombinant overexpression and production of proteins is often done heterologously in *E. coli.* Recombinant metalloproteins are known to be often produced with misincorporated metals or without a metal at all [68-71]. An illustration of this can be seen in a study that measured metal content in recombinant *P. furiosus* proteins expressed in *E. coli* (Table 1.2). Of 45 putative *P. furiosus* ironproteins expressed in *E. coli* only five were shown by ICP-MS to contain iron. Nine of the 45 contained zinc. On the other hand, of the 27 putative *P. furiosus* zinc proteins expressed in *E. coli*, only 6 contained zinc and 6 contained iron. In addition, of 58 *P. furiosus* proteins predicted to contain metals other than iron or zinc, four of those heterologously expressed in *E. coli* contained iron and eight contained zinc. There are 840 hypothetically conserved ORFs in the genome with no prediction for metal content. Of 37 of these heterologously expressed in *E. coli*,

Table 1.2. Predicted versus measured metal contents of recombinant *Pyrococcus furiousus* proteins.

Predicted metal content	# ORFs in *Pf* Genome predicted from InterPro	#*Pf* ORFs expressed in *E. coli*	# proteins containing Fe	Zn
Fe-containing:	306/1352 (23 %)	45	5/45	9/45
Zn-containing:	245/1352 (18 %)	27	6/27	6/27
Not Fe or Zn:	801/1352 (59 %)	58	4/58	8/58
'Unknown' ORFs	840	37	3/37	5/37

three contained iron and five contained zinc. The lack of reliability in metalloprotein predictions shows that certainty of specific metal association with native proteins is most likely determined by analyzing natively purified proteins. Unfortunately, that method is not conducive to genome wide studies since it is labor intensive, time consuming and requires a method of detection required for each enzyme. Moreover, since some metals serve only structural roles, enzyme assays would not aid in the purification of proteins of this type.

Predictive strategies to identify protein complexes.

As mentioned above, a method that is often used to predict protein complexes is based on homology to characterized protein complexes. In addition, a signature sequence such as Cys-X-X-Cys is predictive of metal binding, although it may not indicate what specific metal binds. Metal binding motifs are found in only a small percentage of known metalloproteins and thus are unable to be used to predict metalloproteins lacking those motifs [63; 72-74].

Other strategies that depend on the genome can be helpful in predicting interacting proteins. Operon prediction is useful since cells (especially prokaryotic cells) often maximize efficiency in their response to stimuli by co-expressing proteins that work together in the cell. Because multiple ORFS in an operon are expressed on a polycistronic RNA molecule they are required to be in close proximity to each other in the genome. In fact, algorithms that use the distance between two adjacent genes on the same DNA strand have been used to predict the number of operons in a genome and ultimately the number of protein-protein complexes in prokaryotic genomes [75]. Similarly, transcriptomic analysis can be used to indirectly predict protein-protein complexes by observing which ORFS are co-regulated in response to some environmental challenge. Table 1.3 list ORFs that are dramatically up-regulated in peptide grown cells [50]. Many of those ORFs are adjacent and predicted to form operons. A total of 1460 ORFs are

Table 1.3 ORFs whose expression is up-regulated 5-fold or more in peptide-grown cells and their potential operon arrangement[a]

Function and ORF	Description[c]
PF0289	[Phosphoenolpyruvate carboxykinase]
Aminopeptidase	
PF0366[b]	[Conserved hypothetical protein]
PF0368	[Conserved hypothetical protein]
PF0369	Deblocking aminopeptidase
PF0370	[Phosphoglycerate dehydrogenase]
PF0477	Alpha amylase
[Cobalt metabolism]	
PF0528	[Cobalt transport ATP-binding protein]
PF0529	[Conserved hypothetical protein]
PF0530	[Conserved hypothetical protein]
PF0531	[Cobalamin biosynthesis protein]
Aryl 2-ketoacid metabolism	
PF0532	Acetyl-CoA synthetase II alpha
PF0533	Indolepyruvate Fd oxidoreductase alpha
PF0559	[Hydrogenase regulatory protein]
PF0612	[Conserved hypothetical protein]
PF0613	Fructose-1,6-bisphosphatase
PF0676	Carbamate kinase
PF0689	[Conserved hypothetical protein]
PF0692	[Prismane]
Hydrogenase I	
PF0891	Hydrogenase I beta
PF0892	Hydrogenase I gamma
PF0893	Hydrogenase I delta
PF0894	Hydrogenase I alpha
PF0913	[Conserved hypothetical protein]
PF0915	[Conserved hypothetical protein]
[Acetyl-CoA synthetase]	
PF0972	[Acyl carrier protein synthase]
PF0973	[Acetyl CoA synthase]
PF0974	[Conserved hypothetical protein]
PF1057	[Phophoglycerate kinase]

Table 1.3 continued.

[Amino acid metabolism]

PF1245	[d-Nopaline dehydrogenase]
PF1246	[Sarcosine oxidase, beta]
PF1253	[Aspartate transaminase]
PF1341	[Aminomethyltransferase]

2-Keto acid ferredoxin oxidoreductases

PF1767	2-Ketoglutarate Fd oxidoreductase, delta
PF1768	2-Ketoglutarate Fd oxidoreductase alpha
PF1769	2-Ketoglutarate Fd oxidoreductase beta
PF1770	2-Ketoglutarate Fd oxidoreductase gamma
PF1771	[2-Ketoacid Fd oxidoreductase, alpha]
PF1772	[2-Ketoacid Fd oxidoreductase, beta]
PF1773	[2-Ketoacid Fd oxidoreductase, gamma]
PF1874	[Glyceraldehyde-3-P dehydrogenase]

[Ferredoxin:NADPH oxidoreductase II]

PF1910	[Fd NADPH oxidoreductase II]
PF1911	[Fd NADPH oxidoreductase II]

[Unknown]

PF2001	[Conserved hypothetical protein]
PF2002	[Sucrose transport protein]
PF2047	[1-Asparaginase]

[a]Modified from [50]

[b]Shaded ORFs show genes adjacent in genome.

[c]The ORF description is derived from the genemate annotation (given within brackets) or from experimental data (no brackets).

predicted to be within 470 operons in the *P. furiosus* genome. Of those predicted operons, 349 have been validated by DNA microarray data. A tally of how many of those validated operons encode proteins that form complexes is incomplete. Nonetheless, knowledge of the function of those proteins in the predicted operons is invaluable in helping to focus efforts to identify novel protein-protein interactions.

Methods that detect Protein Complexes.

Genomics has not developed to a point where protein-protein interactions can be predicted from sequence analysis. Protein complexes have historically been identified by purifying an enzyme activity of interest through multiple liquid chromatographic steps followed by analysis of the enzymatically active fractions. Co-elution of two or more polypeptides with enzymatic activity through multiple chromatographic columns is evidence for an enzyme composed of multiple subunits. Furthermore, metal analysis of purified fractions could reveal that an enzyme forms a complex with metal ions.

Protein purification alone is an insufficient strategy to resolve every complex in a cell. Protein purification often requires that function be known before a protein is purified since enzyme activity is routinely used to detect the presence of the protein during purification. Methods have been developed that do not require an explicit knowledge of protein function to make queries regarding protein complex formation [76-78] (Table 1.4). The yeast two hybrid system detects protein-protein interactions by exploiting the modular structure of transcriptional activators [79-81]. Transcriptional activators have a DNA binding domain and a transcriptional activating domain. Hybrid proteins composed of the DNA binding domain and a "bait" protein and the transcriptional activation domain and a "prey" are produced in a cell. If the bait and prey proteins interact they will bring the DNA binding and transcriptional activation domains together

Table 1.4 Comparison of methods to detect protein-protein interactions

Method	Advantages	Disadvantages
Y2H[a] host	Detects transient interactions	Limited to binary interactions; needs genetically tractable
FRET[b] host	Detects transient interactions	Limited to binary interactions; needs genetically tractable
Phage display	Highest Throughput	Not suitable for identifying transient interactions; Limited to binary interactions
TAP[c]	Not limited to binary interactions	Not suitable for identifying transient interactions; needs genetically tractable host

[a] Y2H Yeast two hybrid.
[b] Fluorescence Resonance Energy Transfer
[c] Tandem Affinity Purification

enabling them to express a reporter gene such as β-galactosidase. Yeast two hybrid is generally done in yeast cells which may affect the conformation of the fused bait and prey proteins if they are derived from organisms other than yeast. Furthermore, the fusion itself may affect the behavior of the proteins. Fluorescence resonance energy transfer (FRET) detects the interaction of proteins by tagging interacting proteins with fluorophores [81; 82]. This may be done chemically or genetically. When the donor flouorophore is excited it emits light at a wavelength that is absorbed by the acceptor fluorophore provided the donor and acceptor fluorophores are in close proximity and thus interacting. After the acceptor fluorophore is excited, it emits light at a wavelength which is detected by a fluorometer or with microscopy.

In phage display, cDNAs or segments of genomes are fused with genes that encode phage surface coat proteins [82]. When fusion proteins are expressed on the surface of phages, entire libraries can be screened for interaction between the displayed protein and a protein of interest that is fixed to a surface. After several enrichment steps that wash away viruses that do not bind to the protein of interest, the DNA sequence responsible for the interacting protein is identified by infecting *E. coli* with the phage that displays a protein that specifically interacts with the protein of interest. Infected *E. coli* will produce enough DNA for sequencing and identification of the gene that encodes the interacting protein. As a rule billions of clones can be screened in a week using the phage display technique compared to millions in 2 to 4 weeks with Y2H [83].

Tandem Affinity Purification (TAP) is used to purify stable complexes from whole cells by the use of a tag fused to a protein suspected of having binding partners [84-88]. The tag consists of two affinity domains in tandem. The extreme domain is protein A, followed by a tobacco etch virus (TEV) cleavage site, a calmodulin binding peptide and finally the bait protein. This construct is expressed in the cell where the bait protein is able to interact with its natural

binding partners to form a protein-protein complex. The first purification step involves incubation with IgG beads to pull down the protein complex. This step is followed by cleavage of the TEV cleavage site with TEV protease which exposes the calmodulin binding domain. The second purification step uses calmodulin coated beads in the presence of calcium to pull down the complex from solution. The complex is eluted from the beads with EDTA and the components of the protein complex are identified with mass spectrometry. The advantage of TAP over the yeast two hybrid system, FRET and Phage Display is that it is not limited to binary interactions. The absence of a third protein partner in the binary systems might preclude the interaction of the proteins being tested. The advantage of yeast two hybrid and FRET systems over TAP and Phage Display is that they are able to detect transient interactions whereas in the case of TAP and Phage Display those interactions may be lost during multiple purification steps [89]. In all cases fusion proteins are necessary to carry out the experiments, a feature that may destabilize interactions between protein partners. In addition, execution of these methods are limited to biological systems that have well developed genetic systems.

Recent efforts to decipher the metalloproteome (the set of proteins that have metal-binding capacity) of a given tissue or organism analyzed native cell extracts and side stepped the need for a genetic system [90]. One group used immobilized metal affinity chromatography (IMAC) to analyze the copper and zinc metalloproteomes of human liver cells. The soluble fraction of these cells were loaded onto IMAC columns with either bound copper or zinc. Presumably, proteins that bind copper or zinc would be retained onto the column until they were eluted. Mass spectrometry revealed that different proteins eluted depending on whether the column had bound copper or zinc. Down sides of this method is that it detects mostly abundant metalloproteins, the process only analyzes one metalloproteome at a time and, lastly the method

may not detect metalloproteins that have a metal buried deep within the protein structure. Another methalloproteomic method was used with *Ferroplasma acidifilum* [91]. *F. acidifilum* is an archaeon that lives in heavy metal environments containing high concentrations of iron and sulfur. Its native metalloproteome was analyzed by resolving its proteins on a 2-D native PAGE gel and staining the gel with the chemiluminescent substrate luminal. The protein spots which luminesced because of the presence of a metal were cored and extracted from the gel. Analysis by mass spectrometry and ICPMS identified the protein and its associated metal, respectively. Most proteins found were associated with iron as a structural component as opposed to a functional component due to adaptation of *F. acidifilum* to high concentrations of iron. It is questionable whether this method would work in a proteome that is not adapted to high concentrations of metal since proteins that are not adapted to high metal concentrations may bind less metal and not be as readily detected by this procedure.

Summary and Objectives.

Although the accessibility to whole genome sequences is an essential element for a comprehensive understanding of how organisms function, more information is needed than can be provided by genome sequences alone to meet this aim. Transcriptomics and proteomics can reveal which genes are expressed both temporally and spatially in different cell types and protein structural genomics studies can provide data concerning the structures and folds available which affect the biological function of proteins. These and other available '-omics' technologies complement genomics which in and of itself cannot be used to bridge the gap between the information found in DNA sequences and the cumulative effects of that information on the state of the cell.

The proteome of a cell encompasses all of the proteins that are expressed by a particular genome. The function of many of these proteins has not been discovered. Since the presence of protein complexes are vital to major cellular processes, [92] determining what protein interactions occur in the cell is invaluable in our understanding of the role these proteins fulfill. Initiatives have been put forth by government agencies to facilitate research that would make linkages from the abundance of information found in genomes to the protein interactions that manifest themselves in cells. The objectives of the research discussed in this thesis are:

1. To find and characterize novel protein-protein assemblies in *P furiosus* by analyzing natively fractionated *P. furiosus* biomass.

2. To find novel *P. furiosus* metalloproteins by analyzing natively fractionated *P. furiosus* biomass.

It is understood that the narrow scope of this thesis will not be able to address all the challenges presented by the post-genomic era. However, it is my hope to only to add my small contribution to those of many others with the expectation that more information increases the rate at which the complexities of life can be fashioned into understandable patterns and ultimately be used to manipulate life to the advantage of the human race.

References

1. Alberts, B.M., Botstein, D., Brenner, S., Cantor, C. R., Doolittle, R. F., Hood, L., McKusick, V. A., Nathans, D., Olson, M. V., Orkin, S., Rosenberg, L. E., Ruddle, F. H., Tilghman, S., Tooze, and J. Watson, J. D. 1988. Report of the committee on mapping and sequencing the human genome. National Academy Press, Washington, D.C.

2. Kanehisa, M. 2000. Post-genome Informatics. Oxford University Press, Oxford. 1-23 p.

3. Kyrpides, N.C., "Genomes OnLine Database (GOLD 1.0): a monitor of complete and ongoing genome projects world-wide," *Bioinformatics*, vol. 15, no. 9, pp. 773-4, 1999.

4. Bernal, A., U. Ear and N. Kyrpides, "Genomes OnLine Database (GOLD): a monitor of genome projects world-wide," *Nucleic Acids Res*, vol. 29, no. 1, pp. 126-7, 2001.

5. Liolios, K., N. Tavernarakis, P. Hugenholtz and N.C. Kyrpides, "The Genomes On Line Database (GOLD) v.2: a monitor of genome projects worldwide," *Nucleic Acids Res*, vol. 34, no. Database issue, pp. D332-4, 2006.

6. Liolios, K., K. Mavromatis, N. Tavernarakis and N.C. Kyrpides, "The Genomes On Line Database (GOLD) in 2007: status of genomic and metagenomic projects and their associated metadata," *Nucleic Acids Res*, vol. 36, no. Database issue, pp. D475-9, 2008.

7. Liolios, K., I.M. Chen, K. Mavromatis, N. Tavernarakis, P. Hugenholtz, V.M. Markowitz and N.C. Kyrpides, "The Genomes On Line Database (GOLD) in 2009: status of genomic and metagenomic projects and their associated metadata," *Nucleic Acids Res*, vol. 38, no. Database issue, pp. D346-54, 2010.

8. Pagani, I., K. Liolios, J. Jansson, I.M. Chen, T. Smirnova, B. Nosrat, V.M. Markowitz and N.C. Kyrpides, "The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata," *Nucleic Acids Res*, vol. 40, no. Database issue, pp. D571-9, 2012.

9. Wright, S.H. 2001. Lander celebrates genome milestone in heavily attended talk. *In* MIT Tech Talk. MIT News Office, Cambridge, MA.

10. Riley, M., T. Abe, M.B. Arnaud, M.K. Berlyn, F.R. Blattner, R.R. Chaudhuri, J.D. Glasner, T. Horiuchi, I.M. Keseler, T. Kosuge, H. Mori, N.T. Perna, G. Plunkett, 3rd, K.E. Rudd,

M.H. Serres, G.H. Thomas, N.R. Thomson, D. Wishart and B.L. Wanner, "Escherichia coli K-12: a cooperatively developed annotation snapshot--2005," *Nucleic Acids Res*, vol. 34, no. 1, pp. 1-9, 2006.

11. Keseler, I.M., C. Bonavides-Martinez, J. Collado-Vides, S. Gama-Castro, R.P. Gunsalus, D.A. Johnson, M. Krummenacker, L.M. Nolan, S. Paley, I.T. Paulsen, M. Peralta-Gil, A. Santos-Zavaleta, A.G. Shearer and P.D. Karp, "EcoCyc: a comprehensive view of Escherichia coli biology," *Nucleic Acids Res*, vol. 37, no. Database issue, pp. D464-70, 2009.

12. Christie, K.R., E.L. Hong and J.M. Cherry, "Functional annotations for the Saccharomyces cerevisiae genome: the knowns and the known unknowns," *Trends Microbiol*, vol. 17, no. 7, pp. 286-94, 2009.

13. Griffiths, P.E. and K. Stotz, "Genes in the postgenomic era," *Theor Med Bioeth*, vol. 27, no. 6, pp. 499-521, 2006.

14. Kell, D.B. and S.G. Oliver, "Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era," *Bioessays*, vol. 26, no. 1, pp. 99-105, 2004.

15. Kornberg, A., "Biochemistry matters," *Nat Struct Mol Biol*, vol. 11, no. 6, pp. 493, 2004.

16. Van Regenmortel, M.H., "Reductionism and complexity in molecular biology. Scientists now have the tools to unravel biological and overcome the limitations of reductionism," *EMBO Rep*, vol. 5, no. 11, pp. 1016-20, 2004.

17. Elgar, G. and T. Vavouri, "Tuning in to the signals: noncoding sequence conservation in vertebrate genomes," *Trends Genet*, vol. 24, no. 7, pp. 344-52, 2008.

18. Stetter, K.O., "Hyperthermophilic procaryotes," *FEMS Microbiol Rev*, vol. 18, no. 2-3, pp. 149-158, 1996.

19. Fiala, G. and K.O. Stetter, "Pyrococcus-Furiosus Sp-Nov Represents a Novel Genus of Marine Heterotrophic Archaebacteria Growing Optimally at 100-Degrees C," *Arch Microbiol,* vol. 145, no. 1, pp. 56-61, 1986.

20. Driskill, L.E., K. Kusy, M.W. Bauer and R.M. Kelly, "Relationship between glycosyl hydrolase inventory and growth physiology of the hyperthermophile Pyrococcus furiosus on carbohydrate-based media," *Appl Environ Microb,* vol. 65, no. 3, pp. 893-897, 1999.

21. Schut, G.J., S.D. Brehm, S. Datta and M.W.W. Adams, "Whole-genome DNA microarray analysis of a hyperthermophile and an archaeon: Pyrococcus furiosus grown on carbohydrates or peptides," *Journal of Bacteriology,* vol. 185, no. 13, pp. 3935-3947, 2003.

22. Poole, F.L., B.A. Gerwe, R.C. Hopkins, G.J. Schut, M.V. Weinberg, F.E. Jenney and M.W.W. Adams, "Defining genes in the genome of the hyperthermophilic archaeon *Pyrococcus furiosus*: Implications for all microbial genomes," *J Bacteriol*, vol. 187, no. 21, pp. 7325-7332, 2005.

23. Lipscomb, G.L., K. Stirrett, G.J. Schut, F. Yang, F.E. Jenney, R.A. Scott, M.W.W. Adams and J. Westpheling, "Natural Competence in the Hyperthermophilic Archaeon Pyrococcus furiosus Facilitates Genetic Manipulation: Construction of Markerless Deletions of Genes Encoding the Two Cytoplasmic Hydrogenases," *Appl Environ Microbio,* vol. 77, no. 7, pp. 2232-2238, 2011.

24. Thorgersen, M.P., K. Stirrett, R.A. Scott and M.W.W. Adams, "Mechanism of oxygen detoxification by the surprisingly oxygen-tolerant hyperthermophilic archaeon, Pyrococcus furiosus," *P Natl Acad Sci USA,* vol. 109, no. 45, pp. 18547-18552, 2012.

25. Wang, B.C., M.W. Adams, H. Dailey, L. DeLucas, M. Luo, J. Rose, R. Bunzel, T. Dailey, J. Habel, P. Horanyi, F.E. Jenney, Jr., I. Kataeva, H.S. Lee, S. Li, T. Li, D. Lin, Z.J. Liu,

C.H. Luan, M. Mayer, L. Nagy, M.G. Newton, J. Ng, F.L. Poole, 2nd, A. Shah, C. Shah, F.J. Sugar and H. Xu, "Protein production and crystallization at SECSG -- an overview," *J Struct Funct Genomics*, vol. 6, no. 2-3, pp. 233-43, 2005.

26. Sugar, F.J., F.E. Jenney, Jr., F.L. Poole, 2nd, P.S. Brereton, M. Izumi, C. Shah and M.W. Adams, "Comparison of small- and large-scale expression of selected Pyrococcus furiosus genes as an aid to high-throughput protein production," *J Struct Funct Genomics*, vol. 6, no. 2-3, pp. 149-58, 2005.

27. Jenney, F.E., Jr., P.S. Brereton, M. Izumi, F.L. Poole, 2nd, C. Shah, F.J. Sugar, H.S. Lee and M.W. Adams, "High-throughput production of Pyrococcus furiosus proteins: considerations for metalloproteins," *J Synchrotron Radiat*, vol. 12, no. Pt 1, pp. 8-12, 2005.

28. Robb, F.T., D.L. Maeder, J.R. Brown, J. DiRuggiero, M.D. Stump, R.K. Yeh, R.B. Weiss and D.M. Dunn, "Genomic sequence of hyperthermophile, Pyrococcus furiosus: implications for physiology and enzymology," *Methods Enzymol*, vol. 330, pp. 134-57, 2001.

29. Bergman, N.H., K.D. Passalacqua, P.C. Hanna and Z.S. Qin, "Operon prediction for sequenced bacterial genomes without experimental information," *Appl Environ Microbiol*, vol. 73, no. 3, pp. 846-54, 2007.

30. Moreno-Hagelsieb, G. and J. Collado-Vides, "A powerful non-homology method for the prediction of operons in prokaryotes," *Bioinformatics*, vol. 18 Suppl 1, pp. S329-36, 2002.

31. Salgado, H., G. Moreno-Hagelsieb, T.F. Smith and J. Collado-Vides, "Operons in Escherichia coli: Genomic analyses and predictions," *P Natl Acad Sci USA*, vol. 97, no. 12, pp. 6652-6657, 2000.

32. Verhagen, M.F., A.L. Menon, G.J. Schut and M.W. Adams, "Pyrococcus furiosus: large-scale cultivation and enzyme purification," *Methods Enzymol*, vol. 330, pp. 25-30, 2001.

33. Cody, G.D., N.Z. Boctor, T.R. Filley, R.M. Hazen, J.H. Scott, A. Sharma and H.S. Yoder, Jr., "Primordial carbonylated iron-sulfur compounds and the synthesis of pyruvate," *Science*, vol. 289, no. 5483, pp. 1337-40, 2000.

34. Corliss, D.A., J. A. Baross; and S. E. Hoffmann, "An hypothesis concerning the relationship between submarine hot springs and the origin of life on Earth.," *Oceanlogica Acta SP*, pp. 59-69., 1981.

35. Huber, C. and G. Wachtershauser, "Activated acetic acid by carbon fixation on (Fe,Ni)S under primordial conditions," *Science*, vol. 276, no. 5310, pp. 245-7, 1997.

36. Hashimoto, K., H. Nishi, S. Bryant and A.R. Panchenko, "Caught in self-interaction: evolutionary and functional mechanisms of protein homooligomerization," *Phys Biol*, vol. 8, no. 3, pp. 035007, 2011.

37. Miller, S., A.M. Lesk, J. Janin and C. Chothia, "The accessible surface area and stability of oligomeric proteins," *Nature*, vol. 328, no. 6133, pp. 834-6, 1987.

38. Bamford, D.H., R.J. Gilbert, J.M. Grimes and D.I. Stuart, "Macromolecular assemblies: greater than their parts," *Curr Opin Struct Biol*, vol. 11, no. 1, pp. 107-13, 2001.

39. Wilson, D.N. and K.H. Nierhaus, "Ribosomal proteins in the spotlight," *Crit Rev Biochem Mol Biol*, vol. 40, no. 5, pp. 243-67, 2005.

40. Chiu, W., M.L. Baker and S.C. Almo, "Structural biology of cellular machines," *Trends Cell Biol*, vol. 16, no. 3, pp. 144-50, 2006.

41. Gilbert, H.J., "Cellulosomes: microbial nanomachines that display plasticity in quaternary structure," *Mol Microbiol*, vol. 63, no. 6, pp. 1568-76, 2007.

42. Srere, P.A., "Complexes of sequential metabolic enzymes," *Annu Rev Biochem*, vol. 56, pp. 89-124, 1987.

43. Ovadi, J. and P.A. Srere, "Macromolecular compartmentation and channeling," *Int Rev Cytol*, vol. 192, pp. 255-80, 2000.

44. Ovadi, J. and V. Saks, "On the origin of intracellular compartmentation and organized metabolic systems," *Mol Cell Biochem*, vol. 256-257, no. 1-2, pp. 5-12, 2004.

45. Menon, A.L., F.L. Poole, 2nd, A. Cvetkovic, S.A. Trauger, E. Kalisiak, J.W. Scott, S. Shanmukh, J. Praissman, F.E. Jenney, Jr., W.R. Wikoff, J.V. Apon, G. Siuzdak and M.W. Adams, "Novel multiprotein complexes identified in the hyperthermophilic archaeon Pyrococcus furiosus by non-denaturing fractionation of the native proteome," *Mol Cell Proteomics*, vol. 8, no. 4, pp. 735-751, 2009.

46. Blamey, J.M. and M.W. Adams, "Purification and characterization of pyruvate ferredoxin oxidoreductase from the hyperthermophilic archaeon *Pyrococcus furiosus*," *Biochim Biophy Acta*, vol. 1161, no. 1, pp. 19-27, 1993.

47. Adams, M.W. and A. Kletzin, "Oxidoreductase-type enzymes and redox proteins involved in fermentative metabolisms of hyperthermophilic Archaea," *Adv Protein Chem*, vol. 48, pp. 101-180, 1996.

48. Smith, E.T., J.M. Blamey and M.W. Adams, "Pyruvate ferredoxin oxidoreductases of the hyperthermophilic archaeon, *Pyrococcus furiosus*, and the hyperthermophilic bacterium, *Thermotoga maritima*, have different catalytic mechanisms," *Biochemistry*, vol. 33, no. 4, pp. 1008-1016, 1994.

49. Schut, G.J., A.L. Menon and M.W. Adams, "2-keto acid oxidoreductases from *Pyrococcus furiosus* and *Thermococcus litoralis*," *Method Enzymol*, vol. 331, pp. 144-158, 2001.

50. Schut, G.J., S.D. Brehm, S. Datta and M.W. Adams, "Whole-genome DNA microarray analysis of a hyperthermophile and an archaeon: *Pyrococcus furiosus* grown on carbohydrates or peptides," *J Bacteriol*, vol. 185, no. 13, pp. 3935-3947, 2003.

51. Mai, X. and M.W. Adams, "Purification and characterization of two reversible and ADP-dependent acetyl coenzyme A synthetases from the hyperthermophilic archaeon *Pyrococcus furiosus*," *J Bacteriol*, vol. 178, no. 20, pp. 5897-5903, 1996.

52. Dafforn, T.R., "So how do you know you have a macromolecular complex?," *Acta Crystallogr D Biol Crystallogr*, vol. 63, no. Pt 1, pp. 17-25, 2007.

53. Clarkson, S.M., E.C. Newcomer, E.G. Young and M.W. Adams, "The elemental sulfur-responsive protein (SipA) from the hyperthermophilic archaeon Pyrococcus furiosus is regulated by sulfide in an iron-dependent manner," *J Bacteriol*, vol. 192, no. 21, pp. 5841-3, 2010.

54. Schut, G.J., J. Zhou and M.W. Adams, "DNA microarray analysis of the hyperthermophilic archaeon Pyrococcus furiosus: evidence for anNew type of sulfur-reducing enzyme complex," *J Bacteriol*, vol. 183, no. 24, pp. 7027-36, 2001.

55. Clarkson, S.M., and M.W. Adams, "The hyperthermophilic archaeon P*yrococcus furiosus* utilizes environmental iron(II) monosulfide cluster complexes as an iron source.," *Manuscript in preparation.*, 2013.

56. Lehninger, A.L., "Role of Metal Ions in Enzyme Systems," *Physiol Rev*, vol. 30, no. 3, pp. 393-429, 1950.

57. Bowen, H.J.M. 1966. Trace elements in biochemistry. Academic Press, London.

58. Riordan, J.F., "The role of metals in enzyme activity," *Ann Clin Lab Sci*, vol. 7, no. 2, pp. 119-29, 1977.

59. Seravalli, J. and S.W. Ragsdale, "Expanding the biological periodic table," *Chem Biol*, vol. 17, no. 8, pp. 793-4, 2010.

60. Bertini, I.A.S.H.S.H. 2001. Handbook of metalloproteins. Marcel Dekker, New York.

61. Lippard, S.J.B. 1994. Principles of bioorganic chemistry. University Science Books, Mill Valley, CA, pp 3-41.

62. Castagnetto, J.M., S.W. Hennessy, V.A. Roberts, E.D. Getzoff, J.A. Tainer and M.E. Pique, "MDB: the Metalloprotein Database and Browser at The Scripps Research Institute," *Nucleic Acids Res*, vol. 30, no. 1, pp. 379-82, 2002.

63. Kasampalidis, I.N., I. Pitas and K. Lyroudia, "Conservation of metal-coordinating residues," *Proteins*, vol. 68, no. 1, pp. 123-30, 2007.

64. Andreini, C., I. Bertini, G. Cavallaro, G.L. Holliday and J.M. Thornton, "Metal-MACiE: a database of metals involved in biological catalysis," *Bioinformatics*, vol. 25, no. 16, pp. 2088-9, 2009.

65. Waldron, K.J., J.C. Rutherford, D. Ford and N.J. Robinson, "Metalloproteins and metal sensing," *Nature*, vol. 460, no. 7257, pp. 823-30, 2009.

66. Bork, P., "Powers and pitfalls in sequence analysis: the 70 % hurdle," *Genome Res*, vol. 10, no. 4, pp. 398-400, 2000.

67. Yannone, S.M., S. Hartung, A.L. Menon, M.W. Adams and J.A. Tainer, "Metals in biology: defining metalloproteomes," *Curr Opin Biotechnol*, vol. 23, no. 1, pp. 89-95, 2012.

68. Eidsness, M.K., S.E. O'Dell, D.M. Kurtz, Jr., R.L. Robson and R.A. Scott, "Expression of a synthetic gene coding for the amino acid sequence of Clostridium pasteurianum rubredoxin," *Protein Eng*, vol. 5, no. 4, pp. 367-71, 1992.

69. Meng, L., S. Ruebush, M. D'Souza V, A.J. Copik, S. Tsunasawa and R.C. Holz, "Overexpression and divalent metal binding properties of the methionyl aminopeptidase from Pyrococcus furiosus," *Biochemistry*, vol. 41, no. 23, pp. 7199-208, 2002.

70. Dai, Y., P.C. Wensink and R.H. Abeles, "One protein, two enzymes," *J Biol Chem*, vol. 274, no. 3, pp. 1193-5, 1999.

71. Mori, S., S. Sumino and T. Kasumi, "Substrate specificity of a tripeptidase as a metalloenzyme purified from Lactococcus lactis subsp. lactis biovar. diacetylactis ATCC 13675," *J Biosci Bioeng*, vol. 93, no. 4, pp. 360-6, 2002.

72. Shu, N., T. Zhou and S. Hovmoller, "Prediction of zinc-binding sites in proteins from sequence," *Bioinformatics*, vol. 24, no. 6, pp. 775-82, 2008.

73. Passerini, A., M. Punta, A. Ceroni, B. Rost and P. Frasconi, "Identifying cysteines and histidines in transition-metal-binding sites using support vector machines and neural networks," *Proteins*, vol. 65, no. 2, pp. 305-16, 2006.

74. Andreini, C., I. Bertini and A. Rosato, "A hint to search for metalloproteins in gene banks," *Bioinformatics*, vol. 20, no. 9, pp. 1373-80, 2004.

75. Tran, T.T., P. Dam, Z. Su, F.L. Poole, 2nd, M.W. Adams, G.T. Zhou and Y. Xu, "Operon prediction in Pyrococcus furiosus," *Nucleic Acids Res*, vol. 35, no. 1, pp. 11-20, 2007.

76. Berggard, T., S. Linse and P. James, "Methods for the detection and analysis of protein-protein interactions," *Proteomics*, vol. 7, no. 16, pp. 2833-42, 2007.

77. Piehler, J., "New methodologies for measuring protein interactions in vivo and in vitro," *Curr Opin Struct Biol*, vol. 15, no. 1, pp. 4-14, 2005.

78. Azarkan, M., J. Huet, D. Baeyens-Volant, Y. Looze and G. Vandenbussche, "Affinity chromatography: a useful tool in proteomics studies," *J Chromatogr B Analyt Technol Biomed Life Sci*, vol. 849, no. 1-2, pp. 81-90, 2007.

79. Fields, S. and O. Song, "A novel genetic system to detect protein-protein interactions," *Nature*, vol. 340, no. 6230, pp. 245-6, 1989.

80. Uetz, P. and R.E. Hughes, "Systematic and large-scale two-hybrid screens," *Curr Opin Microbiol*, vol. 3, no. 3, pp. 303-8, 2000.

81. Koegl, M. and P. Uetz, "Improving yeast two-hybrid screening systems," *Brief Funct Genomic Proteomic*, vol. 6, no. 4, pp. 302-12, 2007.

82. Zozulya, S., M. Lioubin, R.J. Hill, C. Abram and M.L. Gishizky, "Mapping signal transduction pathways by phage display," *Nat Biotechnol*, vol. 17, no. 12, pp. 1193-8, 1999.

83. Willats, W.G., "Phage display: practicalities and prospects," *Plant Mol Biol*, vol. 50, no. 6, pp. 837-54, 2002.

84. Bauer, A. and B. Kuster, "Affinity purification-mass spectrometry. Powerful tools for the characterization of protein complexes," *Eur J Biochem*, vol. 270, no. 4, pp. 570-8, 2003.

85. Ho, Y., A. Gruhler, A. Heilbut, G.D. Bader, L. Moore, S.L. Adams, A. Millar, P. Taylor, K. Bennett, K. Boutilier, L. Yang, C. Wolting, I. Donaldson, S. Schandorff, J. Shewnarane, M. Vo, J. Taggart, M. Goudreault, B. Muskat, C. Alfarano, D. Dewar, Z. Lin, K. Michalickova, A.R. Willems, H. Sassi, P.A. Nielsen, K.J. Rasmussen, J.R. Andersen, L.E. Johansen, L.H. Hansen, H. Jespersen, A. Podtelejnikov, E. Nielsen, J. Crawford, V. Poulsen, B.D. Sorensen, J. Matthiesen, R.C. Hendrickson, F. Gleeson, T. Pawson, M.F. Moran, D. Durocher, M. Mann, C.W. Hogue, D. Figeys and M. Tyers, "Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry," *Nature*, vol. 415, no. 6868, pp. 180-3, 2002.

86. Rigaut, G., A. Shevchenko, B. Rutz, M. Wilm, M. Mann and B. Seraphin, "A generic protein purification method for protein complex characterization and proteome exploration," *Nat Biotechnol*, vol. 17, no. 10, pp. 1030-2, 1999.

87. Gavin, A.C., M. Bosche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J.M. Rick, A.M. Michon, C.M. Cruciat, M. Remor, C. Hofert, M. Schelder, M. Brajenovic, H. Ruffner, A. Merino, K. Klein, M. Hudak, D. Dickson, T. Rudi, V. Gnau, A. Bauch, S. Bastuck, B. Huhse, C. Leutwein, M.A. Heurtier, R.R. Copley, A. Edelmann, E. Querfurth, V. Rybin, G. Drewes, M. Raida, T. Bouwmeester, P. Bork, B. Seraphin, B. Kuster, G. Neubauer and G. Superti-Furga, "Functional organization of the yeast proteome by systematic analysis of protein complexes," *Nature*, vol. 415, no. 6868, pp. 141-7, 2002.

88. Puig, O., F. Caspary, G. Rigaut, B. Rutz, E. Bouveret, E. Bragado-Nilsson, M. Wilm and B. Seraphin, "The tandem affinity purification (TAP) method: a general procedure of protein complex purification," *Methods*, vol. 24, no. 3, pp. 218-29, 2001.

89. Collins, M.O. and J.S. Choudhary, "Mapping multiprotein complexes by affinity purification and mass spectrometry," *Curr Opin Biotechnol*, vol. 19, no. 4, pp. 324-30, 2008.

90. She, Y.M., S. Narindrasorasak, S. Yang, N. Spitale, E.A. Roberts and B. Sarkar, "Identification of metal-binding proteins in human hepatoma lines by immobilized metal affinity chromatography and mass spectrometry," *Mol Cell Proteomics*, vol. 2, no. 12, pp. 1306-18, 2003.

91. Ferrer, M., O.V. Golyshina, A. Beloqui, P.N. Golyshin and K.N. Timmis, "The cellular machinery of Ferroplasma acidiphilum is iron-protein-dominated," *Nature*, vol. 445, no. 7123, pp. 91-4, 2007.

92. Alberts, B., "The cell as a collection of protein machines: preparing the next generation of molecular biologists," *Cell*, vol. 92, no. 3, pp. 291-4, 1998.

CHAPTER 2

IDENTIFYING NOVEL PROTEIN COMPLEXES IN THE PROTEOME OF THE HYPERTHERMOPHILIC ARCHAEON *PYROCOCCUS FURIOSUS*

Introduction

Molecular machines like other types of machines are composed of multiple components. All parts are necessary for the machine to function. Furthermore like a blueprint, the genome has the information for the parts of the molecular machines which can be replaced with new parts when old ones are worn out or damaged. Without these molecular machines the processes that govern life would come to a grinding halt. Therefore, it is of immeasurable value to determine the components of these machines and how they work together to carry out their specific functions. A number of methods that determine which components of molecular machines work together have been developed [1-3]. However, limitations exist in their utility due to manipulations at the molecular level that may affect the native conformation of the proteins which are the component parts of the molecular machines. Furthermore, the available methods do not allow for comprehensive analysis, often focusing on one protein-protein interaction at a time.

Stable protein complexes have been historically identified by purification of enzyme activity through multiple chromatographic steps [4; 5]. While this method preserves the native conformation of proteins, it is impractical in broad proteomics analysis of protein complexes because it cannot be used to identify complexes that are not detected by enzymatic assays. Furthermore, identifying protein complexes at a proteomic scale by purifying enzymatic activity

would require various assays to detect various enzymes, a process that could not be done in a timely manner. A procedure developed in the lab of Dr. Michael Adams draws from the well established, non-denaturing method of protein purification and uses a single method of detection, namely mass spectrometry rather than enzyme activity to detect protein-protein interactions.

Of particular interest was the interaction between PF1837 and PF1838 a novel potential complex designated PC-81. This complex is a homolog of two acetyl-CoA synthetases (ACS) that have been purified from *P. furiosus* native biomass. Those complexes form $\alpha_2\beta_2$ heterotetrameric structures with subunits encoded by unlinked genes. Additional ACS subunit homologs are encoded in the genome and all ACS homologs were detected during fractionation by mass spectrometry. This observation served as the basis of recombinant expression experiments performed to examine complex formation of the ACSs. The findings of that study are the subject of chapter 3 of this thesis.

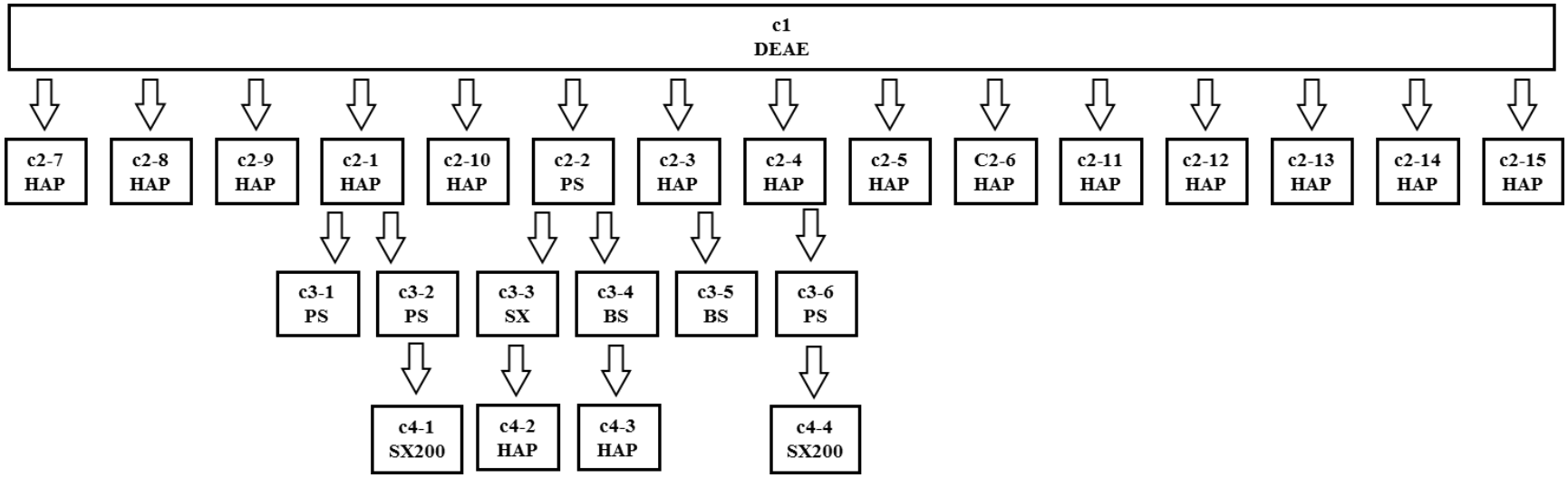
Materials and Methods.

Preparation of anaerobic cell-free extract. *P. furiosus* (DSM 3638) was grown under anaerobic, reducing conditions at 90°C in a 600-liter fermenter on maltose and peptides and harvested in late log phase [6]. The procedures for preparing anaerobic cell-free extract and for running the first chromatography step have been described previously [6] Briefly, 300 g of frozen cells were gently lysed by osmotic shock anaerobically under a continuous flow of argon in 900 ml of 50mM Tris-HCl (pH 8.0) containing 2mM sodium dithionite as a reductant (Buffer A) and 0.5 μg/ml DNAse I to reduce viscosity. The cell lysate was centrifuged at 100,000 x g for 2 h at 18°C, and the clear supernatant, representing the cytoplasmic fraction was immediately loaded onto the first column. This and all subsequent columns were run under anaerobic, reducing

Figure 2.1.

Fractionation schemes of *P. furiosus* cytoplasmic proteins through multiple chromatographic steps.

Cytoplasmic proteins were separated on a DEAE anion exchange column which was the first fractionation. Fractions from the DEAE column were combined into 15 pools that were fractionated through subsequent columns. Some of the fractions of secondary purification were pooled and fractionated further. The chromatography steps included DEAE (diethylaminoethyl cellulose), HAP (hydroxyapatite), PS (phenyl sepharose), BS (blue sepharose), and SX200 (superdex 200) which are indicated in the figure. The fractions generated were analyzed by either CV or HT or both.


c1
DEAE
c2-7
HAP
c2-8
HAP
c2-9
HAP
c2-1
HAP
c2-10
HAP
c2-2
PS
c2-3
HAP
c2-4
HAP
c2-5
HAP
C2-6
HAP
c2-11
HAP
c2-12
HAP
c2-13
HAP
c2-14
HAP
c2-15
HAP
c3-1
PS
c3-2
PS
c3-3
SX
c3-4
BS
c3-5
BS
c3-6
PS
c4-1
SX200
c4-2
HAP
c4-3
HAP
c4-4
SX200

conditions where all buffers were degassed and maintained under a positive pressure of argon, and all liquid transfers were made using needles and syringes.

Large scale anaerobic fractionation of cytoplasmic proteins.. *First column (c1) fractionation.* A column tree representing the complete fractionation is shown in Figure 2.1. All columns were run using an ÄKTA^{TM} basic automated LC system (GE Healthcare). The *P. furiosus* cytoplasmic fraction was diluted 4-fold in Buffer A to reduce the ionic strength and was loaded onto a 10 X 20-cm (1.5-liter) column of DEAE-Sepharose Fast Flow (GE Healthcare) that had been equilibrated previously with the same buffer until the effluent was anaerobic and reducing. Unbound protein was washed off the column with Buffer A. Bound proteins were eluted using a linear gradient (15 liters) of 0-500 mM NaCl in Buffer A, and 125-ml anaerobic fractions were collected (120 fractions). Any remaining tightly bound proteins were eluted from the column using 1-liter step gradients of 1, 1.5, and 2 M NaCl respectively, in Buffer A (high salt washes). *Secondary (c2) chromatographic steps.* A subset of 37 gradient fractions from the first column separation were selected using a combination of protein profiles and enzyme activities to create 6 fraction pools, c1-1 through c1-6. The control proteins and enzyme activities were as follows: c1-1, ferredoxin-NADPH oxidoreductase (FNOR) [7] and rubrerythrin [8]; c1-2, 2-ketisovalerate ferredoxin oxidoreductase (VOR) and pyruvate ferredoxin reductase (POR) [9]; c1-3, aldehyde:ferredoxin oxidoreductase [10]; c1-4, soluble hydrogenase 1 (SH1) [11]; c1-5, rubrdoxin [12]; and c1-6 ferredoxin (Fd)[13]. These pools, c1-1 through c1-6, were fractionated through six subsequent c2 columns. The remaining 83 gradient fractions were combined to generate nine additional pools, in this case based solely on SDS-PAGE and protein profiles. These pools c1-7 through c1-15, were fractionated through nine subsequent c2 columns to give a total of 15 columns at the c2 level (Figure 2.1). SDS-PAGE and native PAGE were carried out

using precast 4-20 % gradient gels (Criterion gel system, Bio-Rad). Protein concentrations were estimated by the method of Bradford [14] using bovine serum albumin as the standard.

Third (c3) and fourth (c4) column steps. Six c3 and four c4 level columns were run using fraction pools that were created for the further purification of the control proteins by the published procedures (Figure 2.1).

Potential protein complex identification using mass spectrometry. Protein digestion, peptide separations, and mass spectrometry were all done in the lab of Gary Siuzdak at the Scripps Research Institute. Protein complex identification was performed computationally using a customized function developed by Farris Poole at the University of Georgia to assess the strength of potential type I complexes based on known type I complexes. Annotation of the proteins are based on the public National Center for Biotechnology Information (NCBI) annotation of the genome [15] except when a protein has been purified from *P. furiosus* or closely related organism or was hypothetical, but contained a domain described in the InterPro database [16]. Details of these methods are given in the original paper.

Results and Discussion

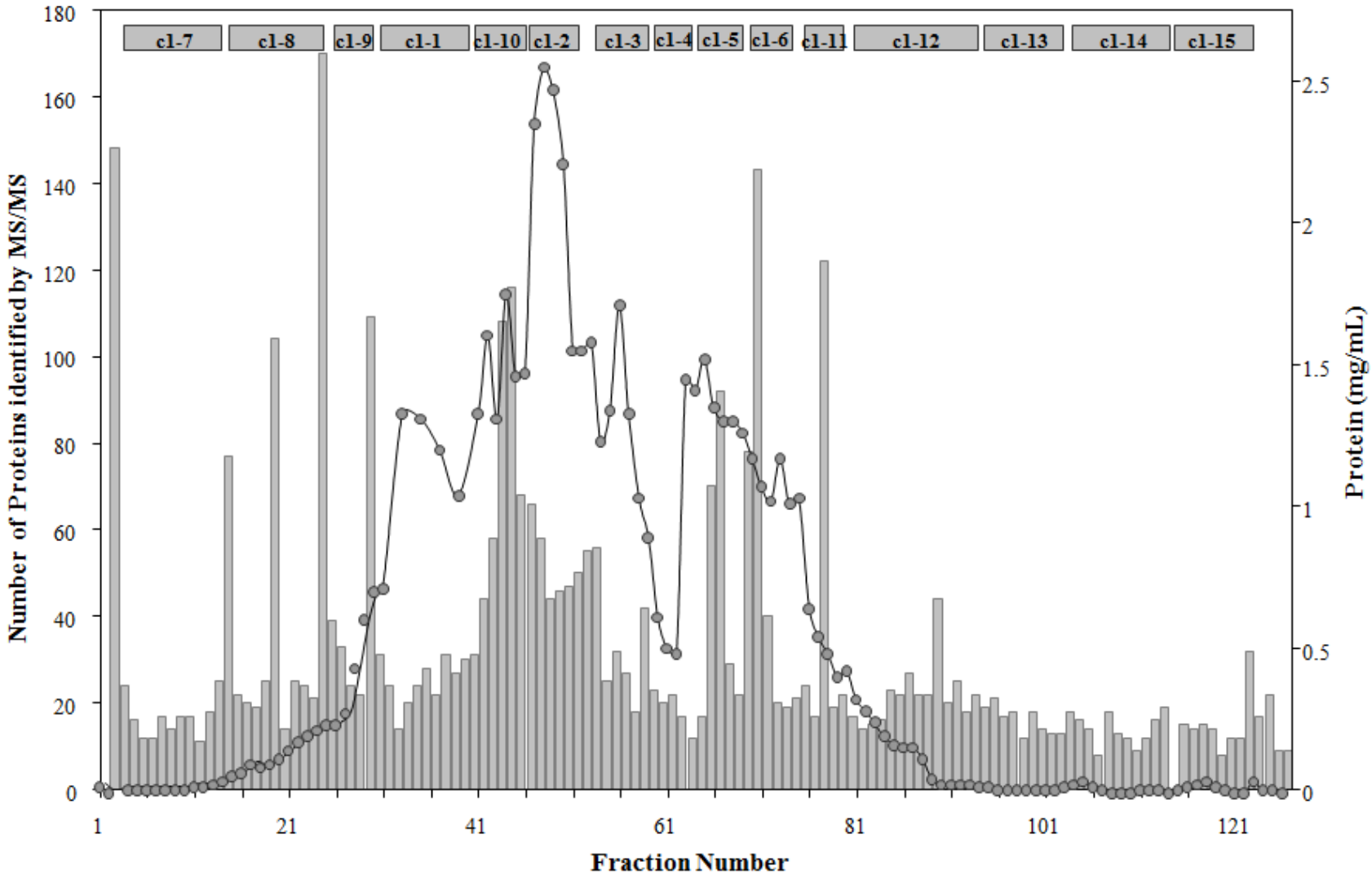
This method was highly dependent upon the sequenced genome of *P. furiosus*. Analysis of the genome of *P. furiosus* which was sequenced in 2001 [17] showed it to be small (1.9 Mb) and well organized with 1,460 of a total of 2,125 total open reading frames (ORFs) predicted to be organized into 470 operons [18]. Genes within in the same operon often have a functional relationship. This relationship may encompass the steps of a biochemical pathway or may represent polypeptides that assemble into a multimeric protein complex. Heteromeric protein complexes of this type are designated type I. Type II complexes are composed of subunits encoded by unlinked genes, and type III complexes are homomeric proteins. (For a discussion of

the different types of multimeric protein complexes see chapter 1). Soluble hydrogenase 1 [11], pyruvate oxidoreductase (POR), 2-ketoisovalerate oxidoreductase (VOR) [9] and ferredoxin $NADP^+$ oxidoreductase (FNOR) [7] are examples of type I complexes since the genes encoding the subunits of these complexes are adjacent in the *P. furiosus* genome. Although, heteromeric proteins encoded by unlinked genes are present in *P. furiosus* [19], we focus here on the purification of complexes encoded by linked genes because access to the genome sequence aids in the systematic prediction of this particular population of novel protein-protein interactions in *P. furiosus*.

Fractionation of native *P. furiosus* biomass was carried out through sequential chromatographic steps which included anion exchange, hydroxyapatite, hydrophobic interaction, and size exclusion chromatography. Since *P. furiosus* grows anaerobically, the fractionation of its biomass was carried out anaerobically to preserve any protein-protein interactions that might be disrupted in the presence of oxygen. 126 column fractions were collected as *P. furiosus* cell free extract was purified over an anion exchange column in the first fractionation. Each fraction was analyzed by MS/MS. Figure 2.2 shows the number of proteins identified in each fraction along with its protein concentrations. Protein identification was determined using conventional (CV) or high through-put (HT) separation techniques followed by tandem mass spectrometry and Mascot searches of the NCBI annotation of the *Pyrococcus furiosus* genome. The CV method requires more time, due to longer HPLC peptide elution gradients, and is not conducive to high-throughput experimentation. However, the CV method has a greater capacity to resolve peptides which increases their detectability by mass spectrometry. Thus, the CV method was used to detect proteins in fractions with greater numbers of species which include those fractions that have been purified by only one chromatographic step.

Figure 2.2

Elution profile of the *P. furiosus* cell free extract fractionation on a DEAE column. Elution profile show number of proteins detected in each fraction (bars), protein concentrations of each fraction (line), and the pools of the DEAE fraction that were further purified.


c1-7
c1-8
c1-9
c1-1
c1-10
c1-2
c1-3
c1-4
c1-5
c1-6
c1-11
c1-12
c1-13
c1-14
c1-15
Number of Proteins identified by MS/MS
Protein (mg/mL)
Fraction Number
0
20
40
60
80
100
120
140
160
180
0
0.5
1
1.5
2
2.5
1
21
41
61
81
101
121

The HT method was typically applied to those fractions that were less complex, since less resolving power is needed when separating a smaller number of peptides. In addition to a shorter time requirement during peptide separations, the HT method further saves time by its adaptation to an automated 96 well format. To test the differences in sensitivity of the CV and HT methods, a set of ten fractions were analyzed by both methods. A total of 403 proteins were detected with387 being identified by the CV method and 111 by the HT method (with 95 detected by both methods). Of those fractions containing more than 60 proteins, the HT method detected less than25 % of those detected by the CV methods. Conversely, the number of proteins detected in fractions that had roughly 20 proteins or less was similar using both approaches.

Identification of type 1 PCs was accomplished by cross tabulation of the identified peptides and the fractions from which they were identified using an Excel pivot table. The identities of the proteins present were listed in the rows and the fractions from which they were detected were listed in the columns. An customized function was written with the pivot data and used to assign an association score based on (i) the co-elution of proteins in common fractions and (ii) if the proteins in common fractions were expressed from adjacent genes. The association scores range from 0.0 defining no association to 1.0 defining the strongest association.

As fractions are purified through sequential orthologous steps, they become less complex and theoretically more enriched with proteins that co-elute due to protein-protein interactions. Those proteins that co-elute through multiple steps of chromatography and were found to be encoded by adjacent genes were considered to form functional interactions. 243 or roughly 25 % of the 967 proteins identified by mass spectrometry were encoded by adjacent genes. These 243 proteins are predicted to form 106 potential type I protein complexes (labeled PC1-PC106). Four previously characterized type I heteromeric enzymes were used as positive controls. They

all had well established purification schemes which were followed and were detectable by enzymatic assays. These included SH1, POR, VOR, and FNOR. Likewise five known homomeric proteins were used as negative controls. Rubredoxin [12] and NADPH-rebredoxin oxidoreductase [20] are monomers, while aldehyde ferredoxin oxidoreductase [10], rubrerythrin [8], and ferredoxin [13] are type III complexes composed of more than one identical subunit.

Known PCs. 61 of the 243 proteins that make up the 106 potential PCs have been previously purified and characterized from *P. furiosus* biomass or in recombinant form. Of the 61, 45 of them are components of 16 potential PCs identified using the method employed in this study (Table 2.1). The proteins that constitute the positive controls are included in this number. Thirteen of the 16 PCs have been purified from *P. furious* native biomass, while three had been characterized using recombinant proteins. Identification of these known PCs validates the criteria used to find novel PCs.

Expected PCs. Five PCs (PC-37 [21], PC-58 [22; 23], PC-63 [24], PC-74 [25], and PC-76 [26]) were categorized as expected because their subunit proteins were homologous to proteins that formed complexes in a closely related organism (Table 2.2). These PCs were involved in protein biosynthesis, RNA processing, central carbon metabolism and amino acid metabolism. With the exception of PC-58, they all have association values between 0.2 and 0.3, and all but PC-58 is predicted to be expressed from an operon by a method used for operon prediction in *P. furiosus* [18].

Novel potential PCs. Eighty-six PCs were deemed novel potential PCs. These were type I PCs that do not fit the definitions of a known or expected PC. In other words, they had not been characterized previously in *P. furiosus* and are neither anticipated due to characterization of homologous complexes in other organisms. The potential PCs have were divided into three

Table 2.1. Known PCs. Dark gray boxes indicate where two or more components of a complex co-elute from a column into the same fraction. Light gray boxes indicate the presence of a subunit that did not co-elute with another component of its complex. Blank squares indicate that the protein was not identified in any fraction. PC-34 represents two known PCs, making the total 16. A.V. is the association value. The protein annotation differs from NCBI data base annotation when the component 'known' as indicated by a reference, is homologous to a known complex in a closely related organism, or has a known InterPro domain (indicated by brackets).

PC	ORF	Annotation	MW	pI	c1	c2-7	c2-8	c2-9	c2-1	c2-10	c2-2	c2-3	c2-4	c2-5	c2-6	c2-11	c2-12	c2-13	c2-14	c2-15	c3-1	c3-2	c3-3	c3-4	c3-5	c3-6	c4-1	c4-2	c4-3	c4-4	A.V.
		CONTROL PCs																													
PC-33	PF0891	**hydrogenase I beta**	43.4	6.3																											0.40
PC-33	PF0892	**hydrogenase I gamma**	33.1	9.0																											0.40
PC-33	PF0893	**hydrogenase I delta**	29.2	5.2																											0.40
PC-33	PF0894	**hydrogenase I alpha**	48.3	5.7																											0.40
PC-34	PF0965	**pyruvate ferredoxin oxidoreductase beta**	36.3	9.1																											0.42
PC-34	PF0966	**pyruvate ferredoxin oxidoreductase alpha**	44.2	5.0																											0.42
PC-34	PF0967	**pyruvate ferredoxin oxidoreductase delta**	12.0	5.2																											0.42
PC-34	PF0968	**2-ketoisovalerate ferredoxin oxidoreductase beta**	34.8	9.1																											0.42
PC-34	PF0969	**2-ketovalerate ferredoxin oxidoreductase alpha**	44.0	5.3																											0.42
PC-34	PF0970	**2-ketisovalerate ferredoxin oxidoreductase delta**	11.9	5.5																											0.42
PC-34	PF0971	**pyruvate/2-ketovalerate ferredoxin oxidoreductase gamma**	20.0	5.4																											0.42
PC-49	PF1327	**ferredoxin:NADPH oxidoreductase (FNOR I alpha)**	52.7	7.1																											0.46
PC-49	PF1328	**ferredoxin:NADPH oxidoreductase (FNOR I beta)**	30.8	7.2																											0.46
		KNOWN PCs																													
PC-2	PF0018	**DNA polymerase II small subunit**	69.2	5.0																											0.20
PC-2	PF0019	**DNA polymerase II large subunit**	143.2	6.2																											0.20
PC-2	PF0020	<beta-lactamase-like (lactamase_B)>	23.8	5.5																											0.20
PC-5	PF0092	**replication factor C, large subunit**	55.3	5.8																											0.26
PC-5	PF0093	**replication factor C, small subunit**	97.8	6.7																											0.26
PC-5	PF0094	**protein disulfide oxidoreductase**	25.8	4.9																											0.26
PC-10	PF0181	[ATPase subunit F]	11.7	5.2																											0.11
PC-10	PF0182	**ATPase subunit A**	113.5	6.1																											0.11
PC-10	PF0183	**ATPase subunit B**	51.9	5.3																											0.11
PC-10	PF0184	**ATPase subunit D**	25.0	9.6																											0.11
PC-22	PF0533	**indolepyruvate oxidoreductase alpha**	71.3	6.8																											0.80
PC-22	PF0534	**indolepyruvate ferredoxin oxidoreductase beta**	23.5	6.4																											0.80
PC-26	PF0598	**aspartate carbamoyltransferase, regulatory subunit**	16.9	7.8																											0.39
PC-26	PF0599	**aspartate carbamoyltransferase, catalytic subunit**	34.7	6.1																											0.39
PC-46	PF1245	**proline dehydrogenase alpha**	54.9	8.1																											0.46
PC-46	PF1246	**proline dehydrogenase beta**	42.5	7.2																											0.46
PC-50	PF1331	**hydrogenase II delta**	26.3	5.9																											0.44
PC-50	PF1332	**hydrogenase II alpha**	46.2	5.0																											0.44
PC-56	PF1433	**mbh11 membrane bound hydrogenase beta**	20.2	5.1																											0.31
PC-56	PF1434	**mbh12 membrane bound hydrogenase alpha**	47.9	6.6																											0.31
PC-62	PF1562	**DNA-directed RNA polymerase subunit a**	44.4	5.7																											0.21
PC-62	PF1563	**DNA-directed RNA polymerase subunit a'**	103.1	6.4																											0.21
PC-62	PF1564	**DNA-directed RNA polymerase subunit b**	127.0	6.3																											0.21
PC-62	PF1565	**DNA-directed RNA polymerase subunit h**	9.2	8.8																											0.21
PC-65	PF1577	hypothetical protein PF1577	48.2	9.3																											0.23
PC-65	PF1578	**DNA topoisomerase VI subunit A**	44.1	8.4																											0.23
PC-65	PF1579	**DNA topoisomerase VI subunit B**	64.4	8.8																											0.23
PC-68	PF1642	**DNA-directed RNA polymerase subunit k**	6.2	9.5																											0.23
PC-68	PF1643	**DNA-directed RNA polymerase subunit n**	8.2	6.1																											0.23
PC-106	PF2018	**replication protein A32**	31.2	5.1																											0.25
PC-106	PF2019	**replication protein A14**	14.2	4.9																											0.25
PC-106	PF2020	**replication protein A41**	41.0	4.9																											0.25

Table 2.2. Expected PCs. For details, see the legend to Table 2.1

PC	ORF	Annotation	MW	pI	c1	c2-7	c2-8	c2-9	c2-1	c2-10	c2-2	c2-3	c2-4	c2-5	c2-6	c2-11	c2-12	c2-13	c2-14	c2-15	c3-1	c3-2	c3-3	c3-4	c3-5	c3-6	c4-1	c4-2	c4-3	c4-4
PC-37	PF0989	phenylalanyl-tRNA synthetase alpha subunit	57.8	9.2																										
PC-37	PF0990	phenylalanyl-tRNA synthetase beta subunit	64.0	4.9																										
PC-58	PF1461	glutamyl-tRNA(Gln) amidotransferase subunit D	48.6	5.7																										
PC-58	PF1462	glutamyl-tRNA(Gln) amidotransferase subunit E	71.2	5.6																										
PC-63	PF1568	[ribonuclease ph (exosome complex)]	27.8	8.9																										
PC-63	PF1569	[RNA binding S1 (exosome complex)]	29.5	8.9																										
PC-74	PF1767	[2-ketoglutarate:ferredoxin oxidoreductase delta]	9.7	6.9																										
PC-74	PF1768	[2-ketoglutarate:ferredoxin oxidoreductase alpha]	44.0	6.1																										
PC-76	PF1795	[sarcosine oxidase, alpha subunit]	51.4	9.1																										
PC-76	PF1796	[polyferredoxin]	18.9	6.9																										
PC-76	PF1797	[L-proline dehydrogenase alpha subunit]	9.8	7.9																										
PC-76	PF1798	[L-proline dehydrogenase beta subunit]	42.4	6.1																										

categories pertaining to the likelihood of their associations. The three categories are weak or unlikely novel PCs, tentative novel PCs, and strong novel potential PCs.

Weak or unlikely novel PCs. Predicted complexes which have one or more of their subunits represented in many fractions, but which are not typically found in the same fraction of its proposed partner are assigned to this category. Thirteen of the potential type I complexes made from a total of 27 proteins are weak or unlikely novel PCs (Table 2.3). An example is PC-4 (PF0042/PF0043), a proposed complex made up of phosphoenolpyruvate synthase [27] and a seemingly unrelated metalloprotease. While phosphoenolpyruvate synthase was found in over150 fractions, its predicted partner was found only in six fractions. Therefore, the co-elution of the proteins encoded by PF0042 and PF0043 of PC-4 is most likely due to coincidence rather than a functional interaction.

Tentative novel PCs. Thirty PCs containing 61 proteins were designated as tentative because they were only identified in fractions of the first purification step (Table 2.4). The validity of 11 of the tentative novel PCs is supported by evidence of them belonging to an operon. These proteins may not have been detected in the later steps of purification because the later steps were analyzed using HT mass spectrometry methodology only, whereas, the first column was analyzed by the CV and HT methods. Some of the complexes detected in the first chromatographic step are likely to be real complexes that are at such a low abundance as not be detected by the HT method in the later columns. However, some potential PCs from the first purification step may have been identified mainly due to the high complexity of early fractions which may have resulted in subunits that were encoded by adjacent genes not being resolved into separate fractions. Strong novel. Forty-three of the potential complexes are characterized as being strongly associated (Table 2.5). Thirty-nine of them are predicted to be encoded by

Table 2.3 Weak or unlikely PCs. For details see legend to table 2.1.

PC	ORF	Annotation	MW	pI	c1	c2-7	c2-8	c2-9	c2-1	c2-10	c2-2	c2-3	c2-4	c2-5	c2-6	c2-11	c2-12	c2-13	c2-14	c2-15	c3-1	c3-2	c3-3	c3-4	c3-5	c3-6	c4-1	c4-2	c4-3	c4-4	A.V.
PC-4	PF0042	metal-dependent hydrolase	28.3	6.0																											0.00
PC-4	PF0043	**phosphoenolpyruvate synthetase**	90.5	5.2																											0.00
PC-13	PF0207	argininosuccinate synthase	46.0	5.3																											0.14
PC-13	PF0208	argininosuccinate lyase	51.3	5.6																											0.14
PC-21	PF0495	**reverse gyrase**	140.0	7.5																											0.18
PC-21	PF0496	<transcriptional regulator winged helix (Wing_hlx_tran_reg)>	30.6	5.6																											0.18
PC-25	PF0588	[phosphoglucomutase]	49.6	5.7																											0.14
PC-25	PF0589	[phosphomannose isomerase]	52.6	5.8																											0.14
PC-31	PF0868	NDP-sugar synthase	47.2	5.1																											0.08
PC-31	PF0869	aspartyl-tRNA synthetase	51.0	5.3																											0.08
PC-44	PF1203	**formaldehyde:ferredoxin oxidoreductase**	69.7	6.0																											0.11
PC-44	PF1204	seryl-tRNA synthetase	53.2	6.7																											0.11
PC-44	PF1205	<nucleotide binding protein, PINc (PINc)>	17.3	8.9																											0.11
PC-67	PF1640	30S ribosomal protein S2	23.0	9.3																											0.07
PC-67	PF1641	<enolase (Enolase)>	37.7	4.4																											0.07
PC-87	PF1931	<circadian clock protein KaiC (KAIC)>	27.7	8.6																											0.13
PC-87	PF1932	<transcriptional regulator winged helix (Wing_hlx_tran_reg)>	31.1	5.2																											0.13
PC-90	PF1993	50S ribosomal protein L10	37.1	4.9																											0.16
PC-90	PF1994	50S ribosomal protein L12	11.2	4.3																											0.16
PC-96	PF0220	hexulose-6-phosphate synthase (d-arabino 3-hexulose 6-phosphate formaldehyde lyase)	47.6	6.3																											0.15
PC-96	PF0221	hypothetical protein PF0221	26.7	5.0																											0.15
PC-97	PF0463	hydrolase related to 2-haloalkanoic acid dehalogenase	27.0	5.1																											0.07
PC-97	PF0464	**glyceraldehyde-3-phosphate ferredoxin oxidoreductase**	73.9	6.3																											0.07
PC-99	PF1499	30S ribosomal protein S19e	17.4	9.7																											0.18
PC-99	PF1500	<PRC-barrel-like (PRCH_cytoplasmic)>	10.4	4.6																											0.18
PC-105	PF1956	**fructose-1,6-bisphosphate aldolase class I (Fba1)**	31.1	6.6																											0.17
PC-105	PF1957	<putative agmatinase (Agmatinase)>	36.0	5.0																											0.17

Table 2.4. Tentative PCs. Tenative complexes were detected only in the DEAE fraction of the first purification step. See Table 1 legend for details.

PC	ORF	Annotation	MW	pI	A.V.
PC-3	PF0037	<armadillo-like helical (ARM-like)>	65.5	8.5	0.19
PC-3	PF0038	<beta-lactamase-like (lactamase_B)>	22.3	6.0	0.19
PC-14	PF0232	transcription regulatory protein, arsR family	21.5	6.2	0.15
PC-14	PF0233	[acetyl-CoA synthetase Q2, alpha subunit]	50.4	6.0	0.15
PC-15	PF0378	50S ribosomal protein L31	11.1	10.1	0.19
PC-15	PF0379	50S ribosomal protein L39e	6.3	12.4	0.19
PC-16	PF0382	prefoldin, beta subunit (pfdA)	19.7	9.5	0.18
PC-16	PF0383	hypothetical protein PF0383	12.8	4.6	0.18
PC-18	PF0431	<phosphoribosylformylglycinamidine cyclo-ligase (purM)>	36.8	5.3	0.23
PC-18	PF0432	<HAD-superfamily hydrolase, subfamily IIA (HAD-SF-IIA)>	30.6	5.2	0.23
PC-19	PF0440	ribonucleotide reductase	200.9	6.1	0.15
PC-19	PF0441	galactose-1-phosphate uridylyltransferase	38.5	8.6	0.15
PC-20	PF0491	<transcription factor TFIIE, alpha subunit (TFIIE_alpha)>	22.9	8.8	0.2
PC-20	PF0492	hypothetical protein PF0492	11.0	5.6	0.2
PC-23	PF0540	<thymidylate synthase complementing protein ThyX (thyX)>	29.0	8.9	0.23
PC-23	PF0541	<Peptidase M24, catalytic core (Peptidase_M24_cat_core)>	32.8	6.7	0.23
PC-27	PF0637	<CRISPR-associated protein, MJ0385 (cas_Csa4)>	39.6	6.3	0.23
PC-27	PF0638	<CRISPR-associated protein, MJ0385 (cas_Csa4)>	44.3	6.8	0.23
PC-36	PF0982	hypothetical protein PF0982	21.6	4.8	0.05
PC-36	PF0983	pcna sliding clamp (proliferating-cell nuclearantigen)	28.0	4.5	0.05
PC-39	PF1115	<PUA (PUA)>	19.5	8.0	0.2
PC-39	PF1116	glycerol-3-phosphate cytidyltransferase	16.9	8.6	0.2
PC-43	PF1189	hypothetical protein PF1189	12.4	5.5	0.14
PC-43	PF1190	<rubrerythrin (Rubrerythrin)>	20.7	5.2	0.14
PC-45	PF1242	dehydrogenase subunit alpha	73.8	8.5	0.23
PC-45	PF1243	hypothetical protein PF1243	11.2	5.0	0.23
PC-53	PF1375	<translation elongation factor EF1A, eukaryotic and archaeal (EF-1_alpha)>	47.6	8.7	0.39
PC-53	PF1376	30S ribosomal protein S10	11.7	10.2	0.39
PC-54	PF1379	cell division protein pelota	40.5	7.2	0.22
PC-54	PF1380	arginyl-tRNA synthetase	72.3	6.2	0.22
PC-55	PF1385	DNA polymerase, bacteriophage-type	22.5	9.0	0.25
PC-55	PF1386	<helix-turn-helix type 3 (HTH_CROC1)>	26.4	9.3	0.25
PC-57	PF1454	<radical SAM (Radical_SAM)>	68.7	6.3	0.2
PC-57	PF1455	<von Willebrand factor, type A (VWFADOMAIN)>	10.3	4.9	0.2
PC-59	PF1482	<beta-grasp fold, ferredoxin-type (Ferredoxin_fold)>	8.2	8.2	0.21
PC-59	PF1483	<DNA-binding SAP (SAP)>	28.3	5.3	0.21

Table 2.4 continued

PC	ORF	Annotation	MW	pI	A.V.
PC-61	PF1556	peptidyl-tRNA hydrolase	13.2	9.5	0.19
PC-61	PF1557	<rossmann-like alpha/beta/alpha sandwich fold (Rossmann-like_a/b/a_fold)>	30.4	5.8	0.19
PC-66	PF1614	hypothetical protein PF1614	57.2	8.6	0.14
PC-66	PF1615	<intein DOD homing endonuclease (INTEIN_ENDONUCLEASE)>	109.7	9.0	0.14
PC-70	PF1716	<nucleotide binding protein, PINc (PINc)>	15.9	9.0	0.19
PC-70	PF1717	translation initiation factor IF-2 gamma subunit	45.0	8.2	0.19
PC-70	PF1718	<wyosine base formation (Wyosine_form)>	39.2	6.9	0.19
PC-71	PF1721	<triphosphoribosyl-dephospho-CoA protein (CitG)>	34.4	7.8	0.15
PC-71	PF1722	archaeal histone a2	7.3	9.6	0.15
PC-73	PF1730	thymidylate kinase	24.2	9.2	0.23
PC-73	PF1731	signal recognition particle SRP54	49.9	9.4	0.23
PC-78	PF1813	50S ribosomal protein L24	14.7	10.4	0.25
PC-78	PF1814	50S ribosomal protein L14	15.2	11.3	0.25
PC-80	PF1827	<uncharacterized conserved protein, 2xCBS, MJ1225 type (UCP_2CBS_MJ1225)>	31.8	8.5	0.18
PC-80	PF1828	hypothetical protein PF1828	14.5	8.9	0.18
PC-84	PF1874	glyceraldehyde-3-phosphate dehydrogenase	37.4	6.0	0.18
PC-84	PF1875	hypothetical protein PF1875	11.4	4.4	0.18
PC-86	PF1926	recombinase (RadA)	38.4	6.2	0.21
PC-86	PF1927	<uncharacterized conserved protein UCP015877, Zn-finger (UCP015877)>	23.5	7.0	0.21
PC-88	PF1974	thermosome, single subunit	60.0	4.9	0.08
PC-88	PF1975	<Phosphoenolpyruvate carboxylase, archaea (DUF557)>	54.3	5.9	0.08
PC-89	PF1990	transcription antitermination protein NusG	16.8	8.2	0.26
PC-89	PF1991	50S ribosomal protein L11	17.6	5.4	0.26
PC-93	PF2039	<transketolase, C-terminal-like>	18.3	5.5	0.23
PC-93	PF2040	<profilin/allergen (PROFILIN)>	20.3	9.1	0.23

Table 2.5. Strong PCs. For details, see the legend in Table 2.1.

PC	ORF	Annotation	MW	pI	c1	c2-7	c2-8	c2-9	c2-1	c2-10	c2-2	c2-3	c2-4	c2-5	c2-6	c2-11	c2-12	c2-13	c2-14	c2-15	c3-1	c3-2	c3-3	c3-4	c3-5	c3-6	c4-1	c4-2	c4-3	c4-4	A.V.
PC-1	PF0014	hypothetical protein PF0014	14.8	6.4																											0.25
PC-1	PF0015	hypothetical protein PF0015	27.1	5.0																											0.25
PC-6	PF0096	<4Fe-4S ferredoxin, iron-sulfur binding (4FE4S_FERREDOXIN)>	8.9	8.2																											0.24
PC-6	PF0097	42 kDa subunit bacteriochlorophyll synthase-like protein	44.1	6.1																											0.24
PC-7	PF0138	<ferritin/ribonucleotide reductase-like (Ferritin/RR_like)>	37.4	5.4																											0.26
PC-7	PF0139	[2-keto acid ferredoxin oxidoreductase 3 gamma]	22.0	8.4																											0.26
PC-8	PF0146	potassium channel, putative	22.1	4.7																											0.31
PC-8	PF0147	potassium channel related protein	23.0	4.8																											0.31
PC-9	PF0160	<peptidase U62, modulator of DNA gyrase (PmbA_TldD)>	50.8	5.1																											0.27
PC-9	PF0161	<peptidase U62, modulator of DNA gyrase (PmbA_TldD)>	46.7	6.7																											0.27
PC-11	PF0198	phosphoribosylformylglycinamidine synthase subunit I	25.4	5.9																											0.24
PC-11	PF0199	phosphoribosylformylglycinamidine synthase subunit II	78.2	5.2																											0.24
PC-12	PF0202	**Isocitrate dehydrogenase**	44.3	6.5																											0.36
PC-12	PF0203	**citrate synthase**	43.0	7.2																											0.36
PC-17	PF0390	hypothetical protein PF0390	42.2	5.3																											1.00
PC-17	PF0391	hypothetical protein PF0391	13.8	8.9																											1.00
PC-24	PF0548	<hydrogenase expression/formation protein (HYPC)>	8.3	4.9																											0.21
PC-24	PF0549	<hydrogenase formation HypD protein (Hydrgn_mat_hypD)>	41.2	6.8																											0.21
PC-28	PF0639	<CRISPR-associated HD (cas3_HD)/CH>	27.1	8.3																											0.25
PC-28	PF0640	<helicase superfamily 1 and 2 ATP-binding (HELICASE_ATP_BIND_1)>	75.3	8.9																											0.25
PC-28	PF0641	<CRISPR-associated protein, MJ0382 (cas_cas5a)>	29.8	9.4																											0.25
PC-29	PF0642	<CRISPR-associated regulatory Csa2 (DUF73)>	37.3	6.4																											0.39
PC-29	PF0643	hypothetical protein PF0643	13.4	6.6																											0.39
PC-30	PF0753	ferrodoxin oxidoreductase beta subunit	32.4	8.8																											0.73
PC-30	PF0754	2-keto acid:ferredoxin oxidoreductase subunit alpha	63.1	5.5																											0.73
PC-32	PF0872	<kaiA binding (KaiA-binding)>	14.7	6.2																											0.23
PC-32	PF0873	hypothetical protein PF0873	5.8	8.1																											0.23
PC-35	PF0972	<hydroxymethylglutaryl-coenzyme A synthase (HMG_coA_syn)>	38.2	7.9																											0.31
PC-35	PF0973	acetyl-CoA acetyltransferase	42.6	4.8																											0.31
PC-35	PF0974	hypothetical protein PF0974	15.9	8.5																											0.31
PC-38	PF1033	<thioredoxin-like fold (IPR012336)>	24.7	5.5																											0.20
PC-38	PF1034	<orotate phosphoribosyl transferase (pyrE)>	20.1	5.5																											0.20
PC-40	PF1121	<CRISPR-associated protein Cas5, Tneap type (cas_Cas5t)>	23.9	9.3																											0.21
PC-40	PF1122	<CRISPR-associated regulatory DevR (cas_Cst2_DevR)>	39.9	6.1																											0.21
PC-40	PF1123	<CRISPR-associated protein, CXXC_CXXC region (cas_CXXC_CXXC)>	55.6	6.7																											0.21
PC-41	PF1124	<CRISPR-associated protein, TM1791 (cas_TM1791_cmr6)>	38.7	8.9																											0.23
PC-41	PF1125	<CRISPR-associated protein, TM1791.1 (cas_Cmr5)>	19.7	8.6																											0.23
PC-41	PF1126	<CRISPR-associated RAMP Cmr4 (cas_RAMP_Cmr4)>	32.6	5.6																											0.23

Table 2.5 continued.

PC	ORF	Annotation	MW	pI	c1	c2-7	c2-8	c2-9	c2-1	c2-10	c2-2	c2-3	c2-4	c2-5	c2-6	c2-11	c2-12	c2-13	c2-14	c2-15	c3-1	c3-2	c3-3	c3-4	c3-5	c3-6	c4-1	c4-2	c4-3	c4-4	A.V.
PC-42	PF1186	<FAD-dependent pyridine nucleotide-disulphide oxidoreductase (Pyr_redox_2)>	48.7	5.7																											0.21
PC-42	PF1187	hypothetical protein PF1187	39.0	5.1																											0.21
PC-47	PF1268	<Cobalamin (vitamin B12)-independent methionine synthase MetE, N-terminal (Meth_synt_1)>	35.5	5.2																											0.23
PC-47	PF1269	methionine synthase	39.3	5.9																											0.23
PC-48	PF1283	**rubrerythrin**	19.5	5.7																											0.16
PC-48	PF1284	<histone-fold (Histone-fold)>	6.8	9.1																											0.16
PC-48	PF1285	<FeS assembly protein SufD (sufD)>	40.2	7.1																											0.16
PC-48	PF1286	<FeS assembly protein SufB (sufB)>	53.5	6.0																											0.16
PC-48	PF1287	<FeS assembly ATPase SufC (SufC)>	27.2	6.1																											0.16
PC-51	PF1339	hypothetical protein PF1339	25.8	5.3																											0.48
PC-51	PF1340	<von Willebrand factor, type A (VWFADOMAIN)>	25.4	8.5																											0.48
PC-52	PF1356	glucose-1-phosphate uridylyltransferase	32.3	5.3																											0.21
PC-52	PF1357	UDP- or dTTP-glucose 4-epimerase or 4-6-dehydratase	38.1	5.8																											0.21
PC-60	PF1529	<vitamin B6 biosynthesis protein (PDXS_SNZ_2)>	36.9	5.8																											0.20
PC-60	PF1530	<thiamine biosynthesis Thi4 protein (Thi4)>	27.2	5.3																											0.20
PC-64	PF1573	<peptidase U62, modulator of DNA gyrase (PmbA_TldD)>	51.1	5.6																											0.25
PC-64	PF1574	<peptidase U62, modulator of DNA gyrase (PmbA_TldD)>	48.8	5.3																											0.25
PC-69	PF1713	carbamoyl-phosphate synthase small subunit	41.7	5.9																											0.34
PC-69	PF1714	carbamoyl-phosphate synthase large subunit	118.1	5.7																											0.34
PC-72	PF1728	NDP-sugar synthase	40.1	5.4																											0.24
PC-72	PF1729	phospho-sugar mutase	26.5	5.6																											0.24
PC-75	PF1771	[2-keto acid:ferredoxin oxidoreductase alpha]	43.1	5.3																											0.40
PC-75	PF1772	[2-keto acid:ferredoxin oxidoreductase beta]	31.1	9.2																											0.40
PC-75	PF1773	[2-keto acid:ferredoxin oxidoreductase gamma]	19.6	6.0																											0.40
PC-77	PF1807	50S ribosomal protein L32e	15.5	11.2																											0.22
PC-77	PF1808	50S ribosomal protein L6	20.9	8.7																											0.22
PC-77	PF1809	30S ribosomal protein S8	14.7	9.6																											0.22
PC-79	PF1824	50S ribosomal protein L4	28.7	10.6																											0.25
PC-79	PF1825	50S ribosomal protein L3	41.4	10.1																											0.25
PC-81	PF1837	**acetyl-CoA synthetase II beta**	26.4	5.6																											0.25
PC-81	PF1838	[acetyl-CoA synthetase NTB alpha]	51.8	5.6																											0.25
PC-82	PF1842	<prokaryotic chromosome segregation and condensation protein ScpA (ScpA_ScpB)>	25.1	4.7																											0.23
PC-82	PF1843	chromosome segregation protein smc	147.8	8.7																											0.23
PC-83	PF1857	cobalt transport ATP-binding protein cbio	24.7	6.4																											0.24
PC-83	PF1858	cysteine synthase	31.6	8.8																											0.24

Table 2.5 continued.

PC	ORF	Annotation	MW	pI	c1	c2-7	c2-8	c2-9	c2-1	c2-10	c2-2	c2-3	c2-4	c2-5	c2-6	c2-11	c2-12	c2-13	c2-14	c2-15	c3-1	c3-2	c3-3	c3-4	c3-5	c3-6	c4-1	c4-2	c4-3	c4-4	A.V.
PC-85	PF1909	**ferredoxin**	7.3	3.9																											0.10
PC-85	PF1910	[sulfide dehydrogenase alpha SudX]	52.5	8.0																											0.10
PC-85	PF1911	[sulfide dehydrogenase SudY]	32.7	5.7																											0.10
PC-91	PF1999	glycine dehydrogenase subunit 1	50.2	5.5																											0.25
PC-91	PF2000	glycine dehydrogenase subunit 2	56.1	6.0																											0.25
PC-92	PF2005	glycerol-3-phosphate dehydrogenase	54.9	5.9																											0.25
PC-92	PF2006	NADH oxidase	45.5	6.3																											0.25
PC-92	PF2007	hypothetical protein PF2007	12.8	9.2																											0.25
PC-94	PF0008	<histidine triad (HIT) protein (HIT_2)>	18.9	6.2																											0.20
PC-94	PF0009	<molybdenum cofactor biosynthesis (MoeB)>	25.5	5.6																											0.20
PC-95	PF0070	<prefoldin (Prefoldin)>	36.9	9.3																											0.22
PC-95	PF0071	<RNA-binding protein, containing Pseudouridine synthase and Archaeosine transglycosylase (PUA) domain>	66.0	8.2																											0.22
PC-98	PF0838	hypothetical protein PF0838	10.0	9.6																											0.22
PC-98	PF0839	<PilT protein, N-terminal (PIN)>	14.7	4.7																											0.22
PC-100	PF1558	30S ribosomal protein S7	24.7	10.0																											0.28
PC-100	PF1559	30S ribosomal protein S12	16.4	10.6																											0.28
PC-101	PF1589	<PfkB (PfkB)>	28.3	6.9																											0.22
PC-101	PF1590	<NUDIX (NUDIX_hydrolase)>	18.9	6.0																											0.22
PC-102	PF1648	30S ribosomal protein S11	14.7	10.4																											0.25
PC-102	PF1649	30S ribosomal protein S4	21.3	10.0																											0.25
PC-102	PF1650	30S ribosomal protein S13	16.9	10.9																											0.25
PC-103	PF1679	<3-isopropylmalate dehydratase large subunit (LeuC)>	41.2	5.5																											0.26
PC-103	PF1680	[3-isopropylmalate dehydratase small subunit]	18.0	6.6																											0.26
PC-104	PF1803	50S ribosomal protein L30	17.7	10.2																											0.21
PC-104	PF1804	30S ribosomal protein S5P	26.6	9.6																											0.21

operons. The components of individual PCs were found to have similar elution profiles, were expressed by genes that appear to be in an operon, were co-regulated at the rRNA level, and were expressed by genes that had similar gene structure to homologous genes found in related organisms. A discussion of several of the strong novel potential PCs found illustrates the potential of this method. A function could not be ascribed to many of the strong PCs because one or more of the components of the complex were describe as "hypothetical" or were poorly annotated or both. PC-17 is an example of a hypothetical complex composed of two hypothetical subunits which are found together in at least 47 fractions indicating that PC-17 is a very stable complex. PC-42 is an example of a hypothetical complex in which only one of its proposed subunits has a known general function. Although it is known that one of the subunits is a flavoprotein, this is not enough information to help discern the function of the whole complex. Similarly, PC-51 has some components that are annotated as homologs of the von Willebrand factor which is found in higher eukaryotes.

Twelve ribosomal proteins form five strong novel PCs (PC-77, PC-79, PC-100, PC-102, and PC-104). These PCs appear to be subcomplexes of complete ribosomes. It may be that the HT mass spectrometry method that detected the complexes did not detect some subunits which were in reality associated with detected subunits. If this is in fact the case, the ribosomal complexes detected are not in reality novel.

Proteins of the clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (cas) system were also detected in the native biomass of *P. furiosus.* CRISPRs are known to be involved in resistance to bacteriophages and constitute a prokaryotic immune system [28]. Eleven of the subunits form four strong novel potential PCs (PC-28, PC-29, PC-40 and PC-41). The genes for PC-28 and PC-29 are predicted to form a

single operon [18] which is down regulated during cold-shock experiments [29]. Indirect evidence from the closely related *T. kodakarensis* using a plasmid challenge assay suggested that PC-28, PC-29, and PC-40 could be involved in CRISPR-Cas mediated defense against foreign DNA in *P. furiosus* [30]. PC-41 contains three of six proteins Cmr [Cas module RAMP (repeat-associated mysterious proteins)] proteins which are important for RNA cleavage [31]. The remaining three which are encoded by PF1128-PF1130 were not identified as a Type I complex. It may be that they do not form a complex or mass spectrometry was unable to detect them.

Based on the NCBI annotation of type I complexes, several strong, novel potential PCs have a role in nucleotide and polynucleotide metabolism. PC-11 is composed of phosphoribosyl-formylglyicinamide synthase subunits I and II. This heterodimeric complex is involved in purine biosynthesis. PC-9 and PC-64 are paralogous complexes and appear to be involved in modulating the activity of DNA gyrase. These two complexes probably account for the unusual DNA topoisomerase and reverse gyrase activites that are unique to hyperthermophilic organisms [32]. The annotations of PC-82 subunits suggest that this complex may be involved in chromosome segregation [33]. This complex was only found in four fractions suggesting it to be in low abundance in the cell. PC-95 forms a complex composed of one subunit that might be involved in RNA metabolism and another one involved in protein folding. The possible relationship between these two functions remains unclear.

Four PCs likely participate in some aspect of carbohydrate metabolism. PC-12 [34; 35] appears to be made of subunits that participate in the tricarboxylic acid cycle, while the others (PC-52, PC-72 and PC-101) are concerned with monosaccharide phosphate derivatives. With the exception of PC-101, each of the carbohydrate metabolism related PCs are predicted to be expressed from an operon [18]. PC-101 is proposed to be composed of a subunit with a

carbohydrate kinase activity and a subunit that is a NUDIX protein. The latter subunit is a type of NDP hydrolase. The crystal structure of a recombinantly produced NUDIX ortholog from *Pyrobaculum aerophilum* has been solved as a homodimeric protein [36]. However, with limited information about the function of these proteins, a physiological heteromeric protein remains a possibility, especially in light of the fact that PC-101 was natively purified.

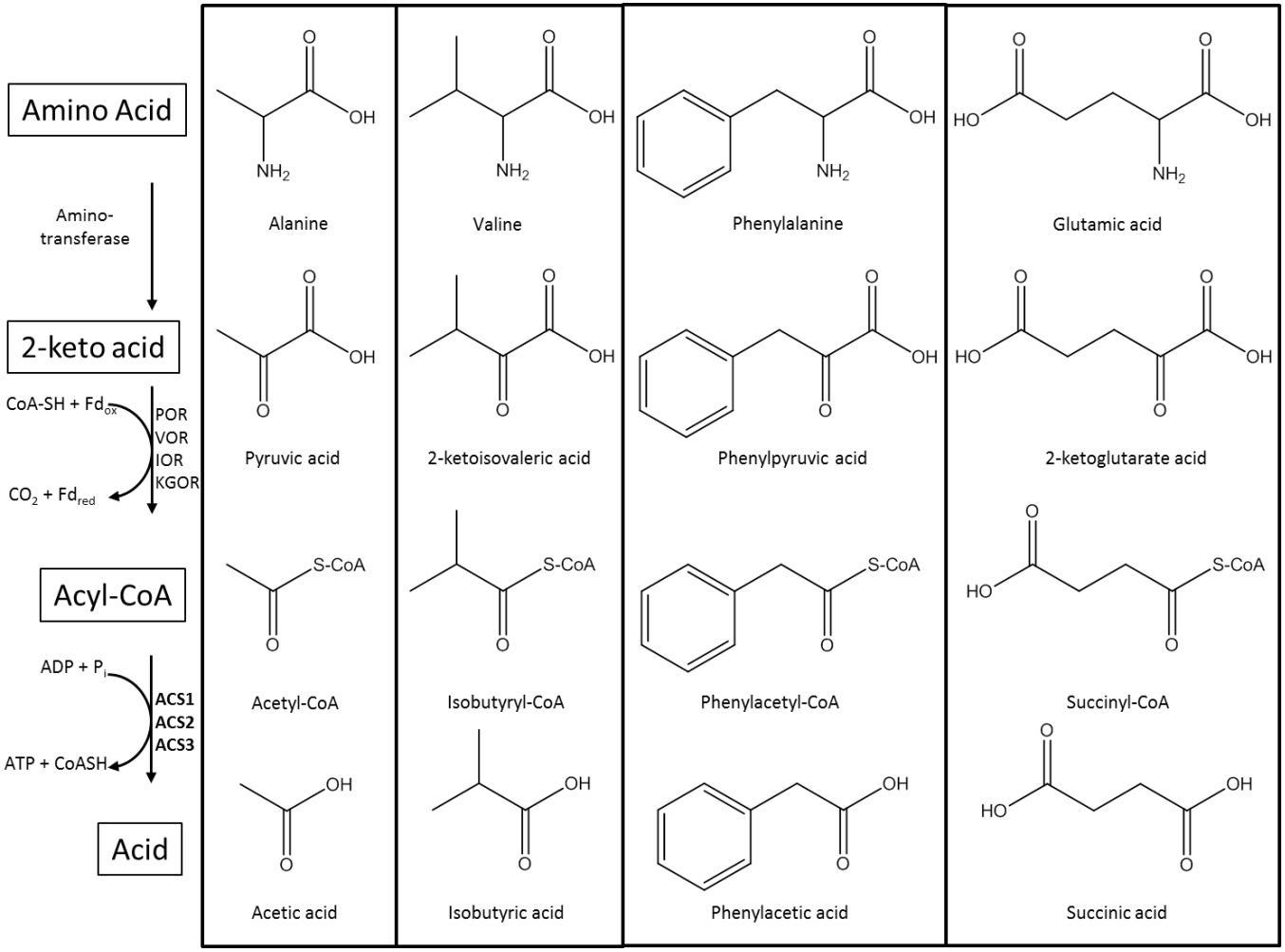
The annotations of proteins in four PCs implicate their involvement in amino acid biosynthesis: PC-47 (methionine), PC-91 (glycine), PC-69 (arginine), and PC-103 (leucine). Experimental evidence for the function of PC-69, which is annotated as a carbamoyl synthetase, shows an up-regulation of the genes that encode these subunits when growing on maltose, an observation which is consistent with a role of this complex in arginine biosynthesis [37]. The annotations of PC-103 (PF1679/PF1680) identify it as a putative 3-isopropylmalate dehydratase. However, PF1679 and PF1680 are paralogs of PF0938 and PF0939 which are in the leucine biosynthesis operon and more likely encode the 3-isopropylmalate dehydratase. A study was done in the closely related *Pyrococcus horikoshii* which suggests that PC-103 is most likely involved lysine rather than leucine biosynthesis [38].

Pyrococcus furiosus can use peptides as a carbon and energy source [39]. Figure 2.3 shows the proposed peptide fermentation pathway. In this pathway, peptides are digested by proteases to individual amino acids. Various 2-ketoacids are produced by the oxidative deamination of the amino acids. This step is carried out by aminotransferases. The resulting 2-ketoacids are converted to acyl-CoA derivatives by ferredoxin (Fd) linked 2-keto acid oxidoreductases (KORs). There are four KORs that have been characterized, each of which has been identified in this study as either a known or expected PC. As mentioned earlier, PC-34 actually represents two type-I complexes, pyruvate ferredoxin: oxidoreductase (POR) [40] and 2-

Figure 2.3.

Proposed pathway of peptide metabolism in fermentative hyperthermophilic archaea.

Fd_{ox}, oxidized ferredoxin; $Fd_{red,}$ reduced ferredoxin; POR, pyruvate oxidoreductase; VOR, 2-ketoisovalerate oxidoreductase; IOR, indolepyruvate oxidoreductase; KGOR, 2-ketoglutarate oxidoreductase; CoA, coenzyme A; ACS acyl-CoA synthetase. ACS1 uses acetyl-CoA and CoA derivatives derived from branched chain amino acids (represented by isobutyric acid in the figure). ACS2 utilizes the substrates of ACS1 in addition to CoA derivatives derived from aromatic amino acids (represented by phenylacetic acid in the figure). ACS3 by homology to succinyl-CoA synthetase of *T. kodakarensis* succinyl-CoA, the CoA derivative derived from Glu/Gln fermentation. POR oxodatively decarboxylates pyruvate. VOR oxidatively decarboxylates branched chain 2-keto acids, IOR oxidatively decarboxylates aromatic 2-keto acids, and KGOR oxidatively decarboxylates 2-ketoglutarate.


Amino Acid
Amino-transferase
2-keto acid
CoA-SH + Fdox
CO2 + Fdred
POR
VOR
IOR
KGOR
Acyl-CoA
ADP + Pi
ATP + CoASH
ACS1
ACS2
ACS3
Acid
Alanine
Valine
Phenylalanine
Glutamic acid
Pyruvic acid
2-ketoisovaleric acid
Phenylpyruvic acid
2-ketoglutarate acid
Acetyl-CoA
Isobutyryl-CoA
Phenylacetyl-CoA
Succinyl-CoA
Acetic acid
Isobutyric acid
Phenylacetic acid
Succinic acid

ketoisovalerate ferredoxin: oxidoreductase (VOR) [41], whose genes are adjacent in the genome (PF0965-PF0971). PC-22 is indolepyruvate ferredoxin: oxidoreductase (IOR) [42] and PC-74 is an ortholog of the 2-ketoglutarate: ferredoxin: oxidoreductase (KGOR) that was characterized in *Thermococcus litoralis* [25]. PC-30 and PC-75 are two additional KOR paralogs that were detected as type-I complexes. Transcriptomic analyses of the genes which encode PC-75 (PF1771-PF1773) show that it is up- regulated in peptide grown cultures while PC-30 (PF0753/PF0754) is not significantly affected [43]. Presumably these KORs encompass the metabolism of additional amino acid derivatives not used by the other KORs since the KORs that have so far been characterized are shown to be involved in the metabolism of at most nine of the 20 amino acids [21; 37; 38].

The reductant formed from the oxidation of 2-ketoacids reduces ferredoxin which in turn reduces $NADP^+$ via ferredoxin: $NADP^+$: oxidoreductase (FNOR I), a known PC (PC-49). PC-85 is categorized as a strong novel type I PC that is a paralog to FNOR I. Called FNOR II, PC-85 is inversely regulated in relation to FNOR1 when grown on peptides and maltose [37]. Although the FNOR homologs catalyze the same reaction, there is likely a physiological significance to their differential regulation. It may be that they affect the regulation of the different KGORs at an enzymatic level during growth on either peptides or carbohydrates.

The acyl-CoA products generated by the KORs are used in ATP synthesis by heterotetrameric ($\alpha_2\beta_2$) NDP- dependent acyl-CoA synthetases. There are a total of five α subunit (PF1540/PF0532/PF0233/PF1085/PF1838) paralogs and two β subunit (PF1787/PF1837) paralogs found in the *P. furiosus* genome. ACS1 (PF1540/PF1787) and ACS2 (PF0532/PF1837) have been purified from native *P. furiosus* biomass [19] and are classified as type II protein complexes because the genes that encode their subunits are unlinked in the

genome. Interestingly, an NDP dependent ACS that was purified and characterized from *Thermococcus kodakarensis* uses succinate as a substrate, indicating its role in glutamate/glutamine metabolism [44]. The *T. kodakarensis* enzyme has the same subunit stoichiometry as the characterized *P. furiosus* ACSs. Its α subunit is an ortholog to PF0233 and its β subunit is an ortholog to the *P. furiosus* ACS2 β (PF1837) subunit. PC-81; (PF1838/PF1837) a strong potential novel type I PC detected by the method used in this study appears to be an additional ACS. Therefore, three ACS enzymes have identical β subunit, but in association with three differing α subunits. The presence of genes for only two β subunits in the genome along with evidence for the association of ACS2 β association with three different ACS α subunits suggest that a variety of combinations of α and β subunits is possible. All five α subunits and both β subunits were detected in the fractions by mass spectrometry although it is difficult to decipher what specific complexes are formed because multiple ACS α and β subunit paralogs co-elute in common fractions most likely due to similar pIs and physical sizes. Similar to the uncharacterized KOR paralogs, characterization of the remaining ACS paralogs will provide a more complete picture of which amino acids participate in the peptide fermentation pathway.

Related methods to the one discussed here have been developed in that they use mass spectrometry to identify intact complexes by co-migration of their components. One such method uses gradient centrifugation to sort high molecular weight complexes in organelles which are then analyzed by mass spectrometry to detect their component parts [45]. A method very similar to the method described here used non-denaturing aerobic fractionation of *E.coli* extract and identifies complexes by co-elution of proteins from chromatographic columns as well [46]. However, they use isobaric tags (iTRAQ) to produce elution profiles for each protein based on

relative abundances to determine which proteins co-elute. While this method is very sophisticated, the investigators only used two levels of purification and only identified abundant and well characterized protein complexes.

Synopsis. A total of 106 potential complexes were identified employing the method described here of identifying co-eluting proteins by mass spectrometry. Eighty-six of the 106 were considered as novel potential and the confidence in their validity as complexes were described as either strong novel, tentative novel, or weak/unlikely novel. Thirty potential PCs were identified in fractions from the first purification step. These complexes were labeled as "tentative" since they were not identified in fractions of downstream columns. Although deemed as such, some "tentative" complexes are likely real and may not have been identified in later fractions because those fractions were analyzed by the HT method rather than the more sensitive CV method. The advantage of the HT method is that it generates data faster than CV method. However, because of the complexity of some fractions, the HT method is unable to detect some proteins that may actually be present in fractions. As a pilot fractionation, a main purpose of this experiment was to see if it was a workable method, therefore, the fractionation was not as extensive as it could have been. Only 5 % of fractions derived from two purification steps were fractionated on additional columns, and only 12 % of those fractions were fractionated further. Clearly, a lot of data is to be had from the fractionation and analysis of every fraction. Furthermore, additional fractionation would reduce the complexity of the fractions and thus make it possible for the HT method to identify a higher percentage of proteins. The added resolution would increase the power of this method by making it possible to not only identify type 1 complexes, but also type 2 complexes as well as low abundant complexes. The low abundance of some complexes, preclude their native characterization. However, once identified, they may be produced recombinantly to

provide enough material for further study. In addition, the data recorded during the execution of this method could be later used as a resource which provides a road map that would be useful in the development of purification schemes for protein complexes of interest.

References

1. Berggard, T., S. Linse and P. James, "Methods for the detection and analysis of protein-protein interactions," *Proteomics*, vol. 7, no. 16, pp. 2833-42, 2007.
2. Piehler, J., "New methodologies for measuring protein interactions in vivo and in vitro," *Curr Opin Struct Biol*, vol. 15, no. 1, pp. 4-14, 2005.
3. Azarkan, M., J. Huet, D. Baeyens-Volant, Y. Looze and G. Vandenbussche, "Affinity chromatography: a useful tool in proteomics studies," *J Chromatogr B Analyt Technol Biomed Life Sci*, vol. 849, no. 1-2, pp. 81-90, 2007.
4. Srere, P.A. and C.K. Mathews, "Purification of Multienzyme Complexes," *Methods in Enzymology*, vol. 182, pp. 539-551, 1990.
5. McHenry, C.S. and W. Crow, "DNA polymerase III of Escherichia coli. Purification and identification of subunits," *J Biol Chem*, vol. 254, no. 5, pp. 1748-53, 1979.
6. Verhagen, M.F., A.L. Menon, G.J. Schut and M.W. Adams, "Pyrococcus furiosus: large-scale cultivation and enzyme purification," *Methods Enzymol*, vol. 330, pp. 25-30, 2001.
7. Ma, K. and M.W. Adams, "Sulfide dehydrogenase from the hyperthermophilic archaeon Pyrococcus furiosus: a new multifunctional enzyme involved in the reduction of elemental sulfur," *J Bacteriol*, vol. 176, no. 21, pp. 6509-17, 1994.
8. Weinberg, M.V., F.E. Jenney, Jr., X. Cui and M.W. Adams, "Rubrerythrin from the hyperthermophilic archaeon Pyrococcus furiosus is a rubredoxin-dependent, iron-containing peroxidase," *J Bacteriol*, vol. 186, no. 23, pp. 7888-95, 2004.

9. Schut, G.J., A.L. Menon and M.W. Adams, "2-keto acid oxidoreductases from *Pyrococcus furiosus* and *Thermococcus litoralis*," *Methods in Enzymology*, vol. 331, pp. 144-158, 2001.

10. Mukund, S. and M.W. Adams, "The novel tungsten-iron-sulfur protein of the hyperthermophilic archaebacterium, Pyrococcus furiosus, is an aldehyde ferredoxin oxidoreductase. Evidence for its participation in a unique glycolytic pathway," *J Biol Chem*, vol. 266, no. 22, pp. 14208-16, 1991.

11. Bryant, F.O. and M.W. Adams, "Characterization of hydrogenase from the hyperthermophilic archaebacterium, Pyrococcus furiosus," *J Biol Chem*, vol. 264, no. 9, pp. 5070-9, 1989.

12. Blake, P.R., J.B. Park, F.O. Bryant, S. Aono, J.K. Magnuson, E. Eccleston, J.B. Howard, M.F. Summers and M.W. Adams, "Determinants of protein hyperthermostability: purification and amino acid sequence of rubredoxin from the hyperthermophilic archaebacterium Pyrococcus furiosus and secondary structure of the zinc adduct by NMR," *Biochemistry*, vol. 30, no. 45, pp. 10885-95, 1991.

13. Aono, S., F.O. Bryant and M.W. Adams, "A novel and remarkably thermostable ferredoxin from the hyperthermophilic archaebacterium Pyrococcus furiosus," *J Bacteriol*, vol. 171, no. 6, pp. 3433-9, 1989.

14. Bradford, M.M., "Rapid and sensitive method for quantitation of microgram quantities of protein utilizing principle of protein-dye binding," *Analytical Biochemistry*, vol. 72, no. 1-2, pp. 248-254, 1976.

15. Pruitt, K.D., T. Tatusova and D.R. Maglott, "NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins," *Nucleic Acids Res*, vol. 35, no. Database issue, pp. D61-5, 2007.

16. Quevillon, E., V. Silventoinen, S. Pillai, N. Harte, N. Mulder, R. Apweiler and R. Lopez, "InterProScan: protein domains identifier," *Nucleic Acids Res*, vol. 33, no. Web Server issue, pp. W116-20, 2005.

17. Robb, F.T., D.L. Maeder, J.R. Brown, J. DiRuggiero, M.D. Stump, R.K. Yeh, R.B. Weiss and D.M. Dunn, "Genomic sequence of hyperthermophile, *Pyrococcus furiosus*: implications for physiology and enzymology," *Methods in Enzymology*, vol. 330, pp. 134-157, 2001.

18. Tran, T.T., P. Dam, Z. Su, F.L. Poole, 2nd, M.W. Adams, G.T. Zhou and Y. Xu, "Operon prediction in Pyrococcus furiosus," *Nucleic Acids Res*, vol. 35, no. 1, pp. 11-20, 2007.

19. Mai, X. and M.W. Adams, "Purification and characterization of two reversible and ADP-dependent acetyl coenzyme A synthetases from the hyperthermophilic archaeon *Pyrococcus furiosus*," *Journal of Bacteriology*, vol. 178, no. 20, pp. 5897-5903, 1996.

20. Ma, K. and M.W. Adams, "A hyperactive NAD(P)H:Rubredoxin oxidoreductase from the hyperthermophilic archaeon Pyrococcus furiosus," *J Bacteriol*, vol. 181, no. 17, pp. 5530-3, 1999.

21. Shiraki, K., M. Tsuji, Y. Hashimoto, K. Fujimoto, S. Fujiwara, M. Takagi and T. Imanaka, "Genetic, enzymatic, and structural analyses of phenylalanyl-tRNA synthetase from Thermococcus kodakaraensis KOD1," *J Biochem*, vol. 134, no. 4, pp. 567-74, 2003.

22. Oshikane, H., K. Sheppard, S. Fukai, Y. Nakamura, R. Ishitani, T. Numata, R.L. Sherrer, L. Feng, E. Schmitt, M. Panvert, S. Blanquet, Y. Mechulam, D. Soll and O. Nureki,

"Structural basis of RNA-dependent recruitment of glutamine to the genetic code," *Science*, vol. 312, no. 5782, pp. 1950-4, 2006.

23. Schmitt, E., M. Panvert, S. Blanquet and Y. Mechulam, "Structural basis for tRNA-dependent amidotransferase function," *Structure*, vol. 13, no. 10, pp. 1421-33, 2005.

24. Ramos, C.R., C.L. Oliveira, I.L. Torriani and C.C. Oliveira, "The Pyrococcus exosome complex: structural and functional characterization," *J Biol Chem*, vol. 281, no. 10, pp. 6751-9, 2006.

25. Mai, X. and M.W. Adams, "Characterization of a fourth type of 2-keto acid-oxidizing enzyme from a hyperthermophilic archaeon: 2-ketoglutarate ferredoxin oxidoreductase from *Thermococcus litoralis*," *Journal of Bacteriology*, vol. 178, no. 20, pp. 5890-5896, 1996.

26. Kawakami, R., H. Sakuraba, H. Tsuge, S. Goda, N. Katunuma and T. Ohshima, "A second novel dye-linked L-proline dehydrogenase complex is present in the hyperthermophilic archaeon Pyrococcus horikoshii OT-3," *FEBS J*, vol. 272, no. 16, pp. 4044-54, 2005.

27. Hutchins, A.M., J.F. Holden and M.W. Adams, "Phosphoenolpyruvate synthetase from the hyperthermophilic archaeon Pyrococcus furiosus," *J Bacteriol*, vol. 183, no. 2, pp. 709-15, 2001.

28. Deveau, H., J.E. Garneau and S. Moineau, "CRISPR/Cas system and its role in phage-bacteria interactions," *Annu Rev Microbiol*, vol. 64, pp. 475-93, 2010.

29. Weinberg, M.V., G.J. Schut, S. Brehm, S. Datta and M.W. Adams, "Cold shock of a hyperthermophilic archaeon: Pyrococcus furiosus exhibits multiple responses to a

suboptimal growth temperature with a key role for membrane-bound glycoproteins," *J Bacteriol*, vol. 187, no. 1, pp. 336-48, 2005.

30. Elmore, J.R., Y. Yokooji, T. Sato, S. Olson, C.V. Glover, 3rd, B.R. Graveley, H. Atomi, R.M. Terns and M.P. Terns, "Programmable plasmid interference by the CRISPR-Cas system in Thermococcus kodakarensis," *RNA Biol*, vol. 10, no. 5, pp. 828-40, 2013.
31. Terns, R.M. and M.P. Terns, "The RNA- and DNA-targeting CRISPR-Cas immune systems of Pyrococcus furiosus," *Biochem Soc Trans*, vol. 41, no. 6, pp. 1416-21, 2013.
32. Forterre, P., "A hot story from comparative genomics: reverse gyrase is the only hyperthermophile-specific protein," *Trends Genet*, vol. 18, no. 5, pp. 236-7, 2002.
33. Tadesse, S. and P.L. Graumann, "Differential and dynamic localization of topoisomerases in Bacillus subtilis," *J Bacteriol*, vol. 188, no. 8, pp. 3002-11, 2006.
34. Steen, I.H., D. Madern, M. Karlstrom, T. Lien, R. Ladenstein and N.K. Birkeland, "Comparison of isocitrate dehydrogenase from three hyperthermophiles reveals differences in thermostability, cofactor specificity, oligomeric state, and phylogenetic affiliation," *J Biol Chem*, vol. 276, no. 47, pp. 43924-31, 2001.
35. Muir, J.M., R.J. Russell, D.W. Hough and M.J. Danson, "Citrate synthase from the hyperthermophilic Archaeon, Pyrococcus furiosus," *Protein Eng*, vol. 8, no. 6, pp. 583-92, 1995.
36. Wang, S., C. Mura, M.R. Sawaya, D. Cascio and D. Eisenberg, "Structure of a Nudix protein from Pyrobaculum aerophilum reveals a dimer with two intersubunit beta-sheets," *Acta Crystallogr D Biol Crystallogr*, vol. 58, no. Pt 4, pp. 571-8, 2002.
37. Schut, G.J., S.D. Brehm, S. Datta and M.W. Adams, "Whole-genome DNA microarray analysis of a hyperthermophile and an archaeon: *Pyrococcus furiosus* grown on

carbohydrates or peptides," *Journal of Bacteriology*, vol. 185, no. 13, pp. 3935-3947, 2003.

38. Lombo, T., N. Takaya, J. Miyazaki, K. Gotoh, M. Nishiyama, T. Kosuge, A. Nakamura and T. Hoshino, "Functional analysis of the small subunit of the putative homoaconitase from Pyrococcus horikoshii in the Thermus lysine biosynthetic pathway," *FEMS Microbiol Lett*, vol. 233, no. 2, pp. 315-24, 2004.

39. Fiala, G. and K.O. Stetter, "*Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100-degrees C," *Archives of Microbiology*, vol. 145, no. 1, pp. 56-61, 1986.

40. Blamey, J.M. and M.W. Adams, "Purification and characterization of pyruvate ferredoxin oxidoreductase from the hyperthermophilic archaeon *Pyrococcus furiosus*," *Biochimica et Biophysica Acta*, vol. 1161, no. 1, pp. 19-27, 1993.

41. Heider, J., X. Mai and M.W. Adams, "Characterization of 2-ketoisovalerate ferredoxin oxidoreductase, a new and reversible coenzyme A dependent enzyme involved in peptide fermentation by hyperthermophilic archaea," *Journal of Bacteriology*, vol. 178, no. 3, pp. 780-787, 1996.

42. Mai, X. and M.W. Adams, "Indolepyruvate ferredoxin oxidoreductase from the hyperthermophilic archaeon *Pyrococcus furiosus*. A new enzyme involved in peptide fermentation," *Journal of Biological Chemistry*, vol. 269, no. 24, pp. 16726-16732, 1994.

43. Schut, G.J., J. Zhou and M.W. Adams, "DNA microarray analysis of the hyperthermophilic archaeon Pyrococcus furiosus: evidence for anNew type of sulfur-reducing enzyme complex," *J Bacteriol*, vol. 183, no. 24, pp. 7027-36, 2001.

44. Shikata, K., T. Fukui, H. Atomi and T. Imanaka, "A novel ADP-forming succinyl-CoA synthetase in *Thermococcus kodakaraensis* structurally related to the archaeal nucleoside diphosphate-forming acetyl-CoA synthetases," *Journal of Biological Chemistry,* vol. 282, no. 37, pp. 26963-26970, 2007.

45. Balbo, A., K.H. Minor, C.A. Velikovsky, R.A. Mariuzza, C.B. Peterson and P. Schuck, "Studying multiprotein complexes by multisignal sedimentation velocity analytical ultracentrifugation," *Proc Natl Acad Sci U S A,* vol. 102, no. 1, pp. 81-6, 2005.

46. Dong, M., L.L. Yang, K. Williams, S.J. Fisher, S.C. Hall, M.D. Biggin, J. Jin and H.E. Witkowska, "A "tagless" strategy for identification of stable protein complexes genome-wide by multidimensional orthogonal chromatographic separation and iTRAQ reagent tracking," *J Proteome Res,* vol. 7, no. 5, pp. 1836-49, 2008.

CHAPTER 3

CHARACTERIZATION OF TEN HETEROTETRAMERIC NDP-DEPENDENT ACYL-COA SYNTHETASES OF THE HYPERTHERMOPHILIC ARCHAEON *PYROCOCCUS FURIOSUS*[1]

[1]Scott, J.W., F.L. Poole, and M.W.W. Adams. 2014. *Archaea*. 2014:Article ID 176863. Reprinted here with permission of the publisher.

Abstract

The hyperthermophilic archaeon *Pyrococcus furiosus* grows by fermenting peptides and carbohydrates to organic acids. In the terminal step, acyl-CoA synthetase (ACS) isoenzymes convert acyl-CoA derivatives to the corresponding acid and conserve energy in the form of ATP. ACS1 and ACS2 were previously purified from *P. furiosus* and have $\alpha_2\beta_2$ structures but the genome contains genes encoding three additional α-subunits. The ten possible combinations of α and β genes were expressed in *E. coli* and each resulted in stable and active $\alpha_2\beta_2$ isoenzymes. The α-subunit of each isoenzyme determined CoA-based substrate specificity and between them they accounted for the CoA derivatives of fourteen amino acids. The β-subunit determined preference for adenine or guanine nucleotides. The GTP-generating isoenzymes are proposed to play a role in gluconeogenesis by producing GTP for GTP-dependent phosphoenolpyruvate carboxykinase and for other GTP-dependent processes. Transcriptional and proteomic data showed that all ten isoenzymes are constitutively expressed indicating that both ATP and GTP are generated from the metabolism of most of the amino acids. A phylogenetic analysis showed that the ACSs of *P. furiosus* and other members of the Thermococcales are evolutionarily distinct from those found throughout the rest of biology, including those of other hyperthermophilic archaea.

Introduction

Members of the Thermococcales order of the Archaea are anaerobic, heterotrophic microorganisms that grow optimally at or above 80°C [1]. They include species of *Pyrococcus* and *Thermococcus*, all of which utilize peptides as carbon and energy sources and many also ferment carbohydrates [2; 3]. Glucose is metabolized via a modified Emden-Meyerhof pathway that involves ADP- rather than ATP-dependent hexo- and phosphofructokinases [4] and a

ferredoxin-linked glyceraldehyde-3-phosphate (GAP) oxidoreductase (GAPOR) rather than an NAD-linked GAP dehydrogenase [5; 6]. The pyruvate that is generated by glycolysis is oxidized by pyruvate ferredoxin oxidoreductase (POR) [7] to yield CO_2 and acetyl-CoA, and energy is conserved in the form of ATP by an ADP-dependent acetyl-CoA synthetase that generates acetate [8]. The general acyl CoA synthetase reaction is shown in (1). This reaction is in contrast to the two-step conversion in anaerobic bacteria whereby acetyl CoA is transformed to acetate via phosphotransacetylase and acetate kinase:

$$R.\,CO.\,SCoA+ADP+Pi \rightarrow RCOOH+ATP+CoASH \quad (1)$$

The proposed pathway for peptide fermentation involves the hydrolysis of peptides into individual amino acids followed by their oxidative deamination to 2-keto acids [9]. Further metabolism of the 2-keto acids is analogous to pyruvate oxidation by POR, in which they are oxidatively decarboxylated to the acyl-CoA derivatives that then serve as substrates for ADP-dependent CoA synthetases. This involves three other ferredoxin-dependent keto acid oxidoreductases. Indolepyruvate ferredoxin oxidoreductase (IOR) and 2-ketoisovalerate ferredoxin oxidoreductase (VOR) oxidize the transaminated products of aromatic amino acids [10] and of branched chain amino acids [11], respectively, while 2-ketoglutarate ferredoxin oxidoreductase (KGOR) oxidizes 2-ketoglutarate, which is derived from glutamate and glutamine [12]. Hence POR is involved in the fermentation of both sugars and peptides (pyruvate is the product of alanine transamination), while IOR, KGOR, and VOR function only in peptide fermentation. The genome of the hyperthermophile *P. furiosus* contains two additional ferredoxin-dependent oxidoreductases. XOR is in an operon with KGOR and is

coregulated with KGOR at the transcriptional level [13] but the substrate specificity of this fifth 2-ketoacid oxidoreductase is not known. In addition, a homolog composed of the gene products of PF0753 and PF0754 was identified as a protein complex during the fractionation of *P. furiosus* biomass [14]. The substrate specificity of this sixth 2-ketoacid oxidoreductase is also unknown.

ACS therefore catalyzes the terminal step of peptide fermentation, the conversion of acyl CoA derivatives to the corresponding acid with concomitant synthesis of ATP [1]. Two forms of ACS, ACS1, and ACS2, have been characterized from *P. furiosus* both in native [8; 15] and recombinant forms [16]. Each consists of two subunits with estimated molecular weights of 45 (α) and 25 kDa (β) with molecular masses for the holoenzymes of approximately 140 kDa, indicative of α2β2 structures. In addition to acetyl CoA, ACS1 used isobutyryl-CoA as a substrate, while ACS2 used phenylacetyl-CoA and indoleacetyl-CoA. It was therefore proposed that ACS1 was involved in the metabolism of alanine and branched chain amino acids and utilized the products of POR and VOR, while ACS2 played a role in the metabolism of aromatic amino acids via IOR [8]. The two enzymes also differed in their specificity for purine nucleotides as ACS1 preferred GDP/GTP to ADP/ATP while ACS2 preferred ADP/ATP to GDP/GTP [8]. From their N-terminal sequences it was evident that ACS1 and ACS2 did not share any subunit [8] and from the *P. furiosus* genome sequence [17] it was evident that ASC1α (PF1540), ACS1β (PF1787), ACS2α (PF0532), and ACS2β (PF1837) are encoded by unlinked genes.

The product of the KGOR reaction is succinyl CoA but no ACS-type activity was detected in extracts of *P. furiosus* cells using this as a substrate [8]. However, an ADP-dependent succinyl-CoA synthetase activity was purified from *Thermococcus kodakarensis*, a

member of the Thermococcales and a close relative of *P. furiosus* [18]. This enzyme, which will be referred to here as ACS3, also had an apparent $\alpha_2\beta_2$ subunit stoichiometry. The gene encoding the α subunit of *T. kodakarensis* ACS3 has a close homolog in the *P. furiosus* genome (PF0233) while that encoding the β-subunit corresponded to PF1837, which encodes the β-subunit of *P. furiosus* ACS2. Hence this β-subunit is associated with two α-subunits to give either *P. furiosus* ACS2 or ACS3. Further inspection of the *P. furiosus* genome reveals that, while there are only two genes that encode ACS β-subunits (PF1837 and PF1787), two additional α-subunits are present encoded by PF1085 and PF1838. These will be designated as ACS4α and ACS5α, respectively. Recently, homologs of these two types of ACSs in *P. furiosus* were characterized from *T. kodakarensis* [19]. It was reported that one of them (TK2127) was specific for 2-(imidazole-4-yl)-acetate, a degradation product of histidine fermentation, while the other (TK0944) had a broad substrate range similar to that of *P. furiosus* ACS2. However, in the *T. kodakarensis* study, TK2127 and TK0944 were characterized with only one of the two possible β-subunits (TK0943) and only using adenine nucleotides.

P. furiosus and other members of the Thermococcales therefore have the genetic potential to synthesize up to ten ACS isoenzymes, depending upon which of the five α-subunits associates with which of the two β-subunits. The corresponding enzymes will be designated herein as ACSx-A or ACSx-G, where A (PF1837) and G (PF1787) represent the two β-subunits encoded by the indicated genes, while *x* (*x* = 1–5) represents the α-subunits encoded by PF1540 (1), PF0532 (2), PF0233 (3), PF1085 (4), and PF1838 (5), respectively. The two β-subunits are 59 % identical at the amino acid sequence level. Hence the two ACS isoforms that have been purified from *P. furiosus* biomass correspond to ACS1-G and ACS2-A, while a third form, ACS3-A, is evident by analogy with the *T. kodakarensis* enzyme. Hence this analysis begs the questions of

whether the α-subunits in these three forms are able to form catalytically active complexes with the alternate β-subunit, which β-subunits are associated with ACS4 and ACS5, and what are the substrate specificities of all of these complexes in terms of both acyl-CoA derivatives and adenine and guanine nucleotides. Herein we have addressed these questions by generating the recombinant forms of the ten possible ACS complexes in *P. furiosus* and assessing their preferences for substrates corresponding to all twenty amino acids.

Materials and methods.

Production of recombinant ACSs. The genes encoding the five α-subunits were amplified by PCR from *P. furiosus* genomic DNA with the following primer sets: PF0233 (ACS3), 5′-AATTTGACATATGACAGTTAACCTAGACTTTC-3′ (+) and 5′-CCGCTCGAGTTAGAGCTCAGCTAAATACTTTC-3′ (−). PF0532 (ACS2), 5′-AATTTGACATATGCTTGACTACTTCTTTAATCCAAG-3′ (+) and 5′-CCGCTCGAGTTAACCATTTACCCCACCTCCAACATTC-3′ (−). PF1085 (ACS4), 5′-AATTTGACATATGAGGTACTTCTTTTACCCAAATAG-3′ (+) and 5′-CCGCTCGAGTCATTCGCTACTAGACCTTAGGCTGAG-3′ (−). PF1540 (ACS1), 5′-AATTTGACATATGAGTTTGGAGGCTCTTTTTAATC-3′ (+) and 5′-CCGCTCGAGTTACTTTTCTTTGTGTTTTGCTTTC-3′ (−). PF1838 (ACS5), 5′-AATTTGACATATGATTAACAACTTGGACATTAAAG-3′ (+) and 5′-CCGCTCGAGTCATTCCCCCATCTTCCTCAAATATTC-3′ (−). NdeI and XhoI restriction sites are underlined for the + and − strands respectively. The resulting DNA fragments encoding the α subunits were each inserted into the pET24a(+) Kan^r expression vector (Novagen, Madison, WI). The genes encoding the two β subunit homologs were also amplified from the *P. furiosus* genomic DNA with the following primers: PF1787 (β-G), 5′-

GATGCCATGGACAGGGTTGCTAAGGCTAGGG-3' (+) and 5′-TAATTTGAGCGGCCGCCTAAAGAATCATCCTAGC-3' (−). PF1837 (β-A), 5′-AATTTGACATATGATTAACAACTTGGACATTAAAG-3' (−). and 5′-CCGCTCGAGTCATTCCCCCATCTTCCTCAAATATTC-3' (−). NcoI and NotI restriction sites are underlined for the + and − strands respectively, for the PF1787 DNA and NdeI and XhoI restriction sites underlined for the + and − strands respectively, for the PF1837 DNA. The amplified PF1787 DNA fragment was inserted into pET21d Amp^r expression vector and the amplified PF1837 DNA fragment was inserted into pET21b Amp^r expression vector. The sequence of each inserted cassette was confirmed by the Sanger method. The α subunit expression plasmids (pET24a(+):PF0233, pET24a(+):PF0532, pET24a(+):PF1085, pET24a(+):PF1540, and pET24a(+):PF1838) were cotransformed with each β-subunit expression plasmid (pET21d:PF1787 and pET21b:PF1837) into *E. coli* BL21(DE3) pRIL.

For recombinant protein production, cultures were grown with the appropriate antibiotics at 37°C in 250 mL 2xYT media to an optical density between 0.8 and 1.0 at 600 nm. Expression of the recombinant genes was induced by the addition of 0.4 mM IPTG (isopropyl-β-D thiogalactopyranoside) at 37°C. Following 3 hours of induction, the cultures were harvested by centrifugation. The pelleted cells were resuspended in lysis buffer (20 mM bis-Tris, pH 7.0, containing 150 mM NaCl, 10 mM MgCl, 1 mM PMSF (phenylmethanesulfonylfluoride), 0.27 mg/mL lysozyme and 0.007 mg/mL DNAse I) and incubated at room temperature for 30 minutes. After centrifugation at 13,000 ×g for 20 min, the supernatant was heated at 80°C for 15 min and centrifuged to remove precipitated heat-labile *E. coli* proteins. The supernatant was applied to a 10 mL HiTrap desalting column (made by connecting two 5 mL columns in tandem) (GE Healthcare) that had been equilibrated with desalting buffer (20 mM bis-Tris, pH 7.0

containing 150 mM NaCl and 10 mM MgCl). The resulting enzyme preparations were used for substrate specificity analyses. Protein concentrations were determined by the Bradford method [20].

To prepare sample for molecular weight determination, the supernatant from the heat treatment step described above (500 μL, ~1.5 mg protein) was diluted 10-fold in buffer A (20 mM bis-Tris, pH 7.0 containing 2 mM dithiothreitol (DTT) and 5 % glycerol) and applied to a 1 mL Mono Q column (GE Healthcare) that had been equilibrated with buffer A. The ACS was eluted with a 20 mL linear gradient from 0 to 0.3 M NaCl in buffer A. Fractions (1 mL) were collected and those containing the ACS as determined by SDS electrophoresis were combined and concentrated to a volume of 200 μL in an Amicon Ultra centrifugation device having a cut-off of 30 kDa. This was applied to a SX200 10/30 (24 mL) gel filtration column that had been previously equilibrated with 50 mM potassium phosphate containing 200 mM KCl, 5 % glycerol, 2 mM DTT, and pH 7.0 and calibrated with standard proteins *β*-amylase (200 kDa), alcohol dehydrogenase (150 kDa), BSA (66 kDa), carbonic anhydrase (29 kDa), and cytochrome c (12.5 kDa).

High-throughput enzyme assays. Since the CoA-forms of the 2-keto acids from all 18 amino acids (assuming that glutamate/glutamine and aspartate/asparagine represent two amino acids) are not commercially available, substrate specificities of each ACS isoenzyme was measured using the reverse reaction of (1) using 18 different acid substrates. A discontinuous assay in which phosphate release from the NTP was quantitated was used in a 96-well format. The ACS sample was that obtained after the heat treatment step described above. The 50 μL reaction mixture contained 20 mM bis-Tris buffer, pH 7.0, 0.5 μg ACS, 20 mM acid substrate, 1 mM ATP or GTP, 2.0 mM $MgCl_2$, and 0.5 mM coenzyme-A. It was heated at 80°C for 1 minute in a 96-

well heating block. The reaction was stopped by adding 1.0 N H_2SO_4. The reaction was found to be linear over at least a 2 min assay period. Free phosphate produced by the hydrolysis of NTP was measured using the BioVision colorimetric assay kit (Mountain View, CA). All liquid handling was carried out robotically (Beckman Coulter Biomek FX, Brea, CA). Kinetic parameters were determined by the same method using NTP substrate concentrations over the range from 0.0625 mM to 4.0 mM. The 96-well plate format was also used to assay activity in the forward direction (1) with the appropriate acyl-CoA derivative. The reaction mixture (50 μL) contained acyl-CoA (0.031 mM to 2.0 mM), K_2HPO_4 (10 mM), $MgCl_2$ (2 mM), ADP (2 mM), and ACS1 (0.5 μg) in 20 mM bis-Tris buffer (pH 7.0). After heating to 80°C for 90 sec, the free CoASH released was measured by the addition of 200 μL of 5′5′-dithiobis(2-nitrobenzoic acid) (DTNB) (0.4 mM) in 100 mM K_2PO_4, pH 7.2.

Results and Discussion

Previously ACS1-G of *P. furiosus* was purified from native biomass [8] and the recombinant forms of the two subunits (ACS1-α and β-G) generated separately in *E. coli* were combined *in vitro* to give a heterotetrameric enzyme that was indistinguishable from that purified from *P. furiosus* [16]. It was unclear, however, if production and combination of individual subunits would be successful for all possible ACS isoenzymes, particularly if cotranslation was required for efficient heteromeric complex formation. Since the *T. kodakarensis* homologs of the *P. furiosus* ACS2-A and ACS3-A isoenzymes (which contain the same β subunit) were produced in *E. coli* using a single plasmid encoding both subunits [18], a similar coexpression approach was used herein. Five plasmids conferring kanamycin resistance were constructed which also encoded each of the five α-subunits, and two plasmids were constructed conferring ampicillin resistance which also encoded each of the β-subunits. Each plasmid encoding a β-

plasmid was then transformed into *E. coli* with a plasmid encoding one of the α-subunits to create all ten possible combinations. Heat treatment of the cell-free extracts of each *E. coli* strain gave rise to two major proteins as determined by SDS PAGE with sizes corresponding to those predicted from the gene sequences of the larger (α, 47.8–51.8 kDa) and smaller (β, 25.8–26.4 kDa) ACS subunits, as shown in Figure 3.1. To determine which of the α- and β-subunit combinations assembled into heteromeric structures, each was analyzed by gel filtration chromatography. All ten eluted with estimated molecular weights of approximately 140 kDa and there was no evidence of individual monomers indicating that all formed stable $\alpha_2\beta_2$ structures when their respective genes were coexpressed in *E. coli*.

The substrate specificities of ten recombinant ACS isoenzymes were determined at 80°C using a 96-well plate approach and the eighteen organic acid substrates that correspond to the twenty amino acids (the same acids are generated by glutamate and glutamine, and by aspartate and asparagine) with either ATP or GTP as the co-substrate. A heat map representing the relative activity of each isoform is shown in Figure 3.2, where 100 % activity represents the highest measured value with either ATP or GTP. A lower limit was arbitrarily set at 30 % of the specific activity produced by the most active substrate and this was used to establish the substrate specificity of each enzyme. As anticipated, ACS1 was most active with acids derived from alanine and the branched chain amino acids, but surprisingly, it also had significant activity with thioglycolate, the end product of cysteine fermentation. In agreement with previous results [8], ACS2 had overlapping specificity with ACS1 activity but, unlike ACS1, it was also highly active with substrates representing the aromatic amino acids. Surprisingly, ACS2 was also active with 3-methylthiopropionate, the end product of methionine fermentation. Hence ACS1 and ACS2 in combination show reasonable activity with acids derived from 9 of the 20 amino

Figure 3.1

SDS-PAGE analyses of recombinant *P. furiosus* ACS isoenzymes.

The lanes correspond to the following: 1: ACS1-G; 2: ACS1-A; 3: ACS2-G; 4: ACS2-A; 5: ACS3-G; 6: ACS3-A, 7: ACS4-G; 8: ACS4-A; 9: ACS5-G; and 10: ACS5-A. Each lane contains approximately 2 μg of the protein obtained after the heat-treatment step. Lanes M indicate molecular weight markers (given in kDa).

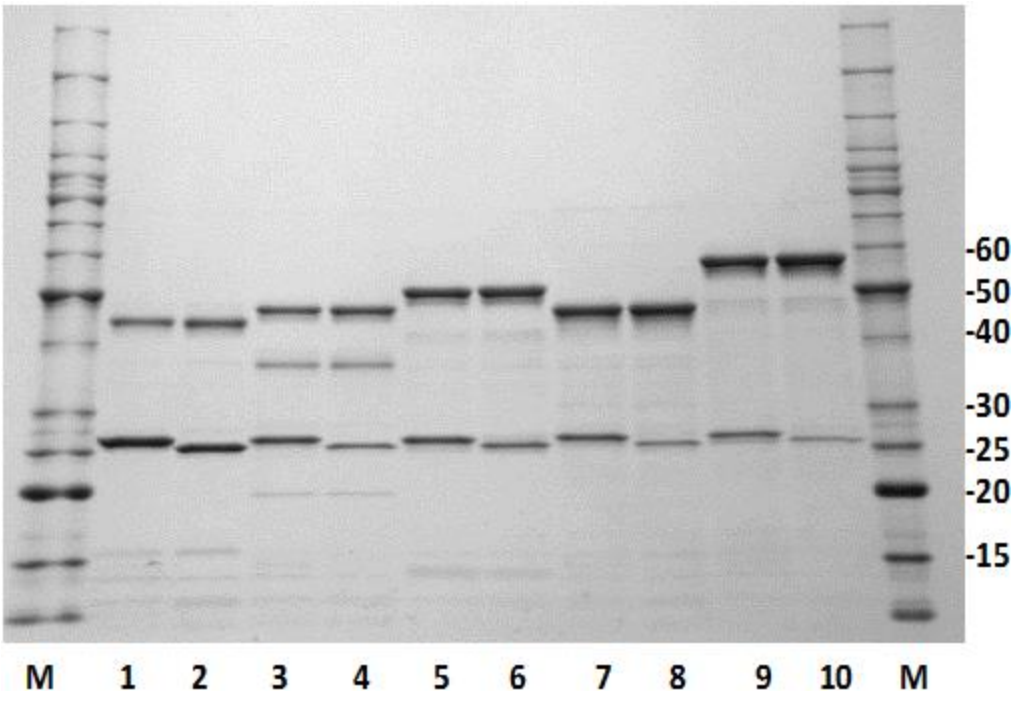
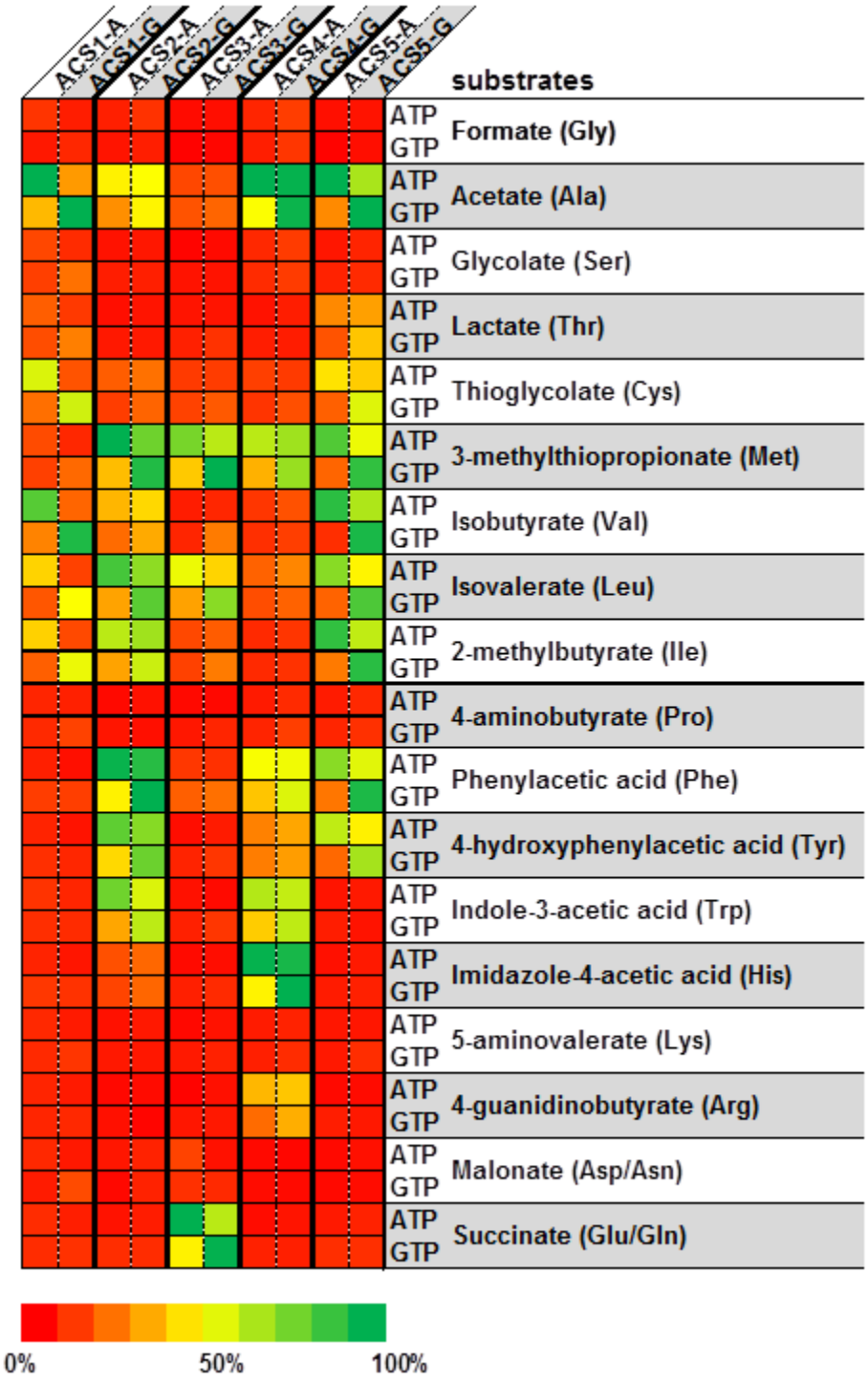

-60
-50
-40
-30
-25
-20
-15
M
1
2
3
4
5
6
7
8
9
10
M

Figure 3.2

Heat map of activities of the ten ACS isoenzymes with eighteen organic acids with ATP or GTP as cosubstrates.

100 % activity corresponds to the highest activity of an individual enzyme with ATP or GTP. All other activities in a given column are relative to the highest value.


ACS1-A
ACS1-G
ACS2-A
ACS2-G
ACS3-A
ACS3-G
ACS4-A
ACS4-G
ACS5-A
ACS5-G
substrates
ATP
GTP
Formate (Gly)
Acetate (Ala)
Glycolate (Ser)
Lactate (Thr)
Thioglycolate (Cys)
3-methylthiopropionate (Met)
Isobutyrate (Val)
Isovalerate (Leu)
2-methylbutyrate (Ile)
4-aminobutyrate (Pro)
Phenylacetic acid (Phe)
4-hydroxyphenylacetic acid (Tyr)
Indole-3-acetic acid (Trp)
Imidazole-4-acetic acid (His)
5-aminovalerate (Lys)
4-guanidinobutyrate (Arg)
Malonate (Asp/Asn)
Succinate (Glu/Gln)
0%
50%
100%

acids. One of the exceptions is succinate, derived from glutamate and glutamine via 2-ketoglutarate. The only isoenzyme that showed activity with succinate was ACS3 (Figure 3.2). However, ACS3 had similar activity with the methionine derivative, comparable to that of ACS2.

ACS4 was the only isoenzyme that shows significant activity with imidazole 4-acetate, which is derived from histidine. It is also the only ACS that shows detectable activity with the organic acid derived from arginine (Figure 3.2). On the other hand, like ACS2, the substrate specificity of ACS4 indicates that this isoenzyme is involved in the degradation of alanine and the aromatic amino acids. ACS5 also shows a broad substrate range similar to that of ACS2, and similarly does not utilize derivatives of methionine or lysine. However, ACS5 is the only enzyme that has significant activity with lactate, which is the end product of the threonine pathway. Considering all ten isoenzymes, there are six amino acids whose acid derivatives werenot utilized by any of them and these were glycine, serine, proline, lysine, aspartate, and asparagine. Of the fourteen acids utilized by the recombinant ACSs in the work presented herein, nine of them (acetate, 3-methylthiopropionate, isobutyrate, isovalerate, 2-methylbutyrate, phenylacetic acid, 4-hydroxyphenylacetic acid, indole-3-acetic acid, and succinate) result from the four KORs that have been characterized [10-12]. However, the genome of *P. furiosus* harbors two additional KORs. The genes for one (XOR, PF1771-PF1773) are expressed and are regulated by carbon source [13] while the genes for the other (PF0753, PF0754) are also expressed and the corresponding proteins have been identified in *P. furiosus* biomass [14]. Furthermore, it is not known if the KORs that have been characterized can use as substrates the harbors two additional KORs. The genes for one (XOR, PF1771-PF1773) are expressed and are regulated by carbon source [13] while the genes for the other (PF0753, PF0754) are also

expressed and the corresponding proteins have been identified in *P. furiosus* biomass [14]. Furthermore, it is not known if the KORs that have been characterized can use as substrates the harbors two additional KORs. The genes for one (XOR, PF1771-PF1773) are expressed and are regulated by carbon source [13] while the genes for the other (PF0753, PF0754) are also expressed and the corresponding proteins have been identified in *P. furiosus* biomass [14]. Furthermore, it is not known if the KORs that have been characterized can use as substrates the activities described herein for *P. furiosus* and also for *T. kodakarensis* [19]. On the other hand, the biosynthetic pathways for amino acids in *P. furiosus* appear to be generally similar to the well-studied routes in mesophilic bacteria. For example, DNA microarray analyses were consistent with the biosynthesis of twelve amino acids (Glu, Arg, Leu, Val, Ile, Ser, Thr, Met, His, Phe, Trp, and Tyr) occurring by conventional pathways [13]. However, growth studies to determine amino acid auxotrophy have given ambiguous results [21-23] and while the genome of *P. furiosus* does encode genes that could be involved in the biosynthesis of all twenty amino acids, some pathways for amino acids may be incomplete [24]. A more comprehensive understanding of amino acid metabolism in *P. furiosus* in general is clearly needed before a more definitive correlation between biosynthetic and degradative amino acid pathways can be made and the corresponding roles of the ACS isoenzymes.

Using the 96-well plate approach, K_m values for ATP and acetate of 0.34 and 1.6 mM, respectively, were determined at 80°C for recombinant ACS1-G. These values are similar to those reported for the same ACS isoenzyme purified from *P. furiosus* native biomass [8], values that were obtained using the same assay conditions as used herein but with one at a time measurements. However, the values are at least two-fold higher than those previously reported for the recombinant ACS-G isoenzyme (0.090 and 0.8 mM, resp. [25], although those data were

based on a coupled assay at 55°C, which might account for the differences. From the data presented in Figure 3.2 it is evident that all five ACS α-subunits utilize acetyl CoA (forward reaction of (1)) and the kinetic constants for each ACS-A enzyme is given in Table 3.1. The K_m and the k_{cat}/K_m values are quite similar for ACS1, ACS2, ACS4 and ACS5, with ACS3 having comparable affinity but much lower activity. Hence all four isoenzymes likely utilize acetyl CoA *in vivo*. Similarly, the results indicate that ACS2 and ACS5 are comparably active with CoA derivatives derived from branched chain amino acids (Table 3.1). Unfortunately, the primary substrate for ACS3, succinyl-CoA, could not be used at 80°C as it is extremely unstable at the assay temperature and direct comparisons were not possible. Similarly, the primary CoA substrates for ACS4 and ACS5, derived from histidine and threonine, are not commercially available. It is interesting to compare the results described herein with *P. furiosus* to those recently reported with *T. kodakaraensis* by Shikata et al. [18] and Awano et al. [19]. For example, while the substrate specificities for the ACS3 homolog from each organism were very similar, *P. furiosus* ACS4 utilized acids derived from histidine, methionine, phenylalanine, tryptophan and arginine, while its counterpart in *T. kodakaraensis* was only reactive with the histidine derivative. Similarly, in *P. furiosus* the substrate range of ACS5 is extended compared to that of the *T. kodakaraensis* homolog as it was also active with lactate, a derivative of threonine fermentation. While kinetic comparisons are limited since the *T. kodakaraensis* homologs were studied mainly using acids as the substrates, in both studies ACS3 exhibited Michaelis-Menten kinetics when acetyl CoA and ADP were used as substrates. However, the *T. kodakaraensis* enzyme had a much lower K_m value (41 μM) and higher catalytic efficiency (0.460 $\mu M^{-1} \cdot s^{-1}$) than *P. furiosus* ACS3 (474 μM and 0.012 $\mu M^{-1} \cdot s^{-1}$). It is not clear to what extent this reflects physiological differences between the two organisms. The specificities for

Table 3.1. Kinetic parameters of ACSx-A isoforms with CoA derivatives.

Enzyme	Substrate	K_m (μM)	k_{cat}/K_m ($\mu M^{-1} \cdot s^{-1}$)
ACS1-A	acetyl-CoA	137 ± 91	0.140
ACS2-A	acetyl-CoA	350 ± 2	0.114
ACS2-A	isovaleryl-CoA	466 ± 119	0.109
ACS2-A	phenylacetyl-CoA	3320 ± 1320	0.150
ACS3-A	acetyl-CoA	474 ± 165	0.012
ACS4-A	acetyl-CoA	242 ± 46	0.119
ACS5-A	acetyl-CoA	620 ± 1.0	0.112
ACS5-A	isobutyryl-CoA	2730 ± 1080	0.083
ACS5-A	isovaleryl-CoA	820 ± 407	0.120

the organic substrates of the ACS isoenzymes of *P. furiosus* containing each of the five α-subunits was independent of whether they were associated with the β-A or β-G subunits (Figure 3.2). However, the type of β subunit did determine the nucleotide specificity. With all ten isoenzymes, those containing the β-A subunit were more active with ATP than GTP. In contrast, enzymes containing the β-G subunit were generally more active with GTP than ATP or at least showed comparable activity with ATP and GTP (Figure 3.2). This was further demonstrated by the kinetic parameters of ACS1-G and ACS1-A with GTP and ATP as substrates.

The catalytic efficiency of ACS1-A was approximately 14-fold higher with ATP than with GTP, while the opposite was true with ACS1-G (Table 3.2). Hence the nucleotide specificity of the ACS isoenzymes is determined by the β-subunit. This is consistent with the presence of two ATP-grasp fold subdomains in the β-subunits (residues 23–139 in Aβ and 25–139 in Gβ), one of which includes a conserved histidinyl residue (His71) at the active site [26]. *P. furiosus* ACSs presumably transfer the phosphate group from the α-subunit to this conserved histidine of the G-β subunit, which then transfers the phosphate group to the nucleoside diphosphate [26].

Hence, the ten ACS isoenzymes in *P. furiosus* can account for the metabolism of the majority (14 of 20) of amino acids, using both adenine and guanine nucleotides. The ACS reaction is the only means of conserving energy during peptide degradation, and the same is true for sugar fermentation. This is due to the unusual modification of the glycolytic pathway in *P. furiosus,* where GAP oxidation is coupled to the reduction of ferredoxin instead of NAD, resulting in the conservation of energy at the ferredoxin oxidation step via a respiratory

Table 3.2 Nucleotide kinetics of ACS1 isoforms.

Isoform	Nucleotide	K_m (μM)	k_{cat}/K_m ($\mu M^{-1} \cdot s^{-1}$)
ACS1-A	ATP	132 ± 19	0.14
ACS1-A	GTP	223 ± 58	0.01
ACS1-G	ATP	342 ± 63	0.011
ACS1-G	GTP	87 ± 7	0.144

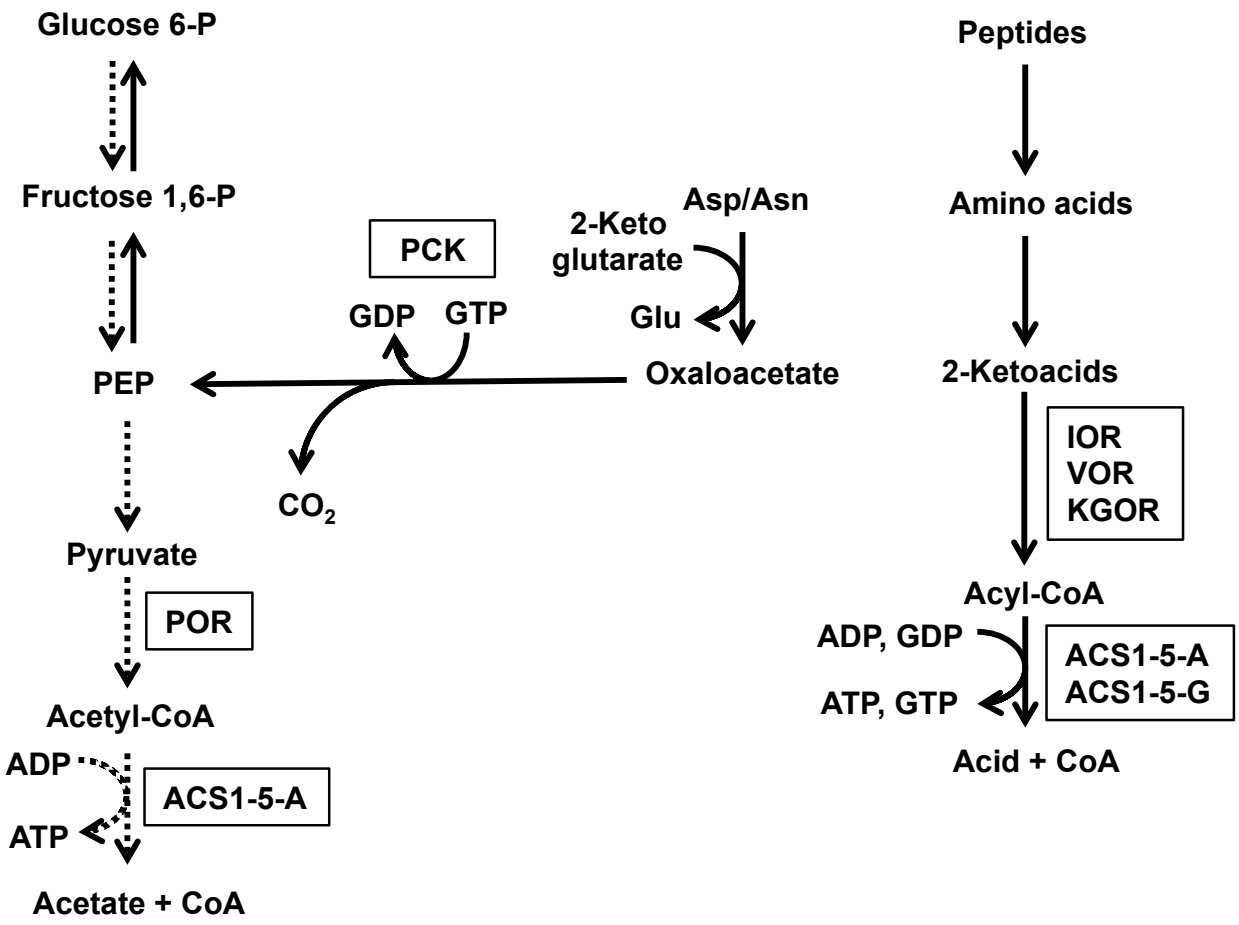
membrane-bound hydrogenase [27; 28] rather than by substrate level phosphorylation (SLP) via 1,3-bisphosphoglycerate [3; 29]. Hence the ACS reaction generating acetate is the only source of ATP via SLP. The results presented herein show that all five ACS-A isoenzymes appear to be efficient in catalyzing this reaction, as well as being involved in peptide degradation. Since ATP is the principle energy carrier within the cell, it would be expected that all ACS-A isoenzymes (containing the β-A subunit) would catalyze these reactions. It is therefore not obvious what advantage is gained by using the β-G subunit to produce GTP. It would seem futile for an NDP kinase to transphosphorylate ADP with the GTP since ATP is made via ACS isoenzymes with the β-A subunit.

Insight into the function of ACS isoenzymes containing the β-G subunit is provided by Fukuda and coworkers [30] who described a GTP-dependent phosphoenolpyruvate carboxykinase (PCK) from *T. kodakarensis*. PCKs have been historically classified into GTP- and ATP-dependent enzymes, with ATP-dependent enzymes described in bacteria, yeast, plants, and some eukaryotic parasites while the GTP-dependent enzyme being found in mammals and some other eukaryotes [31]. Interestingly, transcription of the gene encoding PCK was upregulated when *T. kodakarensis* cells were grown under gluconeogenic (with pyruvate and peptides as the carbon sources) rather than under glycolytic conditions (using starch as the carbon source). DNA microarray analysis showed that the same is true in *P. furiosus* when cells are grown on peptides and a sugar (maltose) [13]. As shown in Figure 3.3, PCK catalyzes the first step of gluconeogenesis, the decarboxylation of oxaloacetate to yield phosphoenolpyruvate and CO_2. Hence we propose that during growth of *P. furiosus* by peptide fermentation, the function of the five ACS-G isoenzymes containing the β-G subunit is to provide GTP for the

Figure 3.3

Proposed function of ACS β-G and ACS β-A isoenzymes during growth of *P. furiosus* on carbohydrates or peptides.

The abbreviations are: POR, pyruvate ferredoxin oxidoreductase; KAOR, 2-keto acid oxidoreductases (IOR, VOR, KGOR and POR, see text for details).


Glucose 6-P
Fructose 1,6-P
PEP
Pyruvate
POR
Acetyl-CoA
ADP
ATP
ACS1-5-A
Acetate + CoA
PCK
GDP
GTP
CO2
Asp/Asn
2-Keto glutarate
Glu
Oxaloacetate
Peptides
Amino acids
2-Ketoacids
IOR
VOR
KGOR
Acyl-CoA
ADP, GDP
ATP, GTP
ACS1-5-A
ACS1-5-G
Acid + CoA

gluconeogenic-related PCK reaction. The ATP-independent generation of GTP by the ACS-G isoenzymes could also provide GTP for the archaeal translational machinery [32].

These data therefore suggest that all ten isoenzymes have physiological relevance in *P. furiosus*. That they all are functional is supported by DNA microarray data, which show that all ten genes are expressed under both gluconeogenic and glycolytic growth conditions and there appears to be little if any differential expression between the two growth modes for any of these genes [13]. The one exception is that the gene encoding the α-subunit of ACS2 (PF0532), which forms an operon with the genes encoding IOR, is up-regulated during growth on peptides. Proteomic analyses of *P. furiosus* grown on maltose and peptides also showed that all ten subunits of the ACS isoenyzmes are detected during fractionation of a cytoplasmic extract [14]. Moreover, immunoprecipitation analyses of *T. kodakarensis* showed that the β-G and the β-A subunits associated with multiple α-subunits [18]. Taken together, these data show that the ten ACS isoenzymes are constitutively produced and are able to generate both ATP and GTP from the wide variety of coenzyme-A derivatives produced during peptide fermentation. The same appears to be true for all members of the Thermococcales since, like *P. furiosus,* their genomes also contain genes encoding two β-subunits and five α-subunits of ACS.

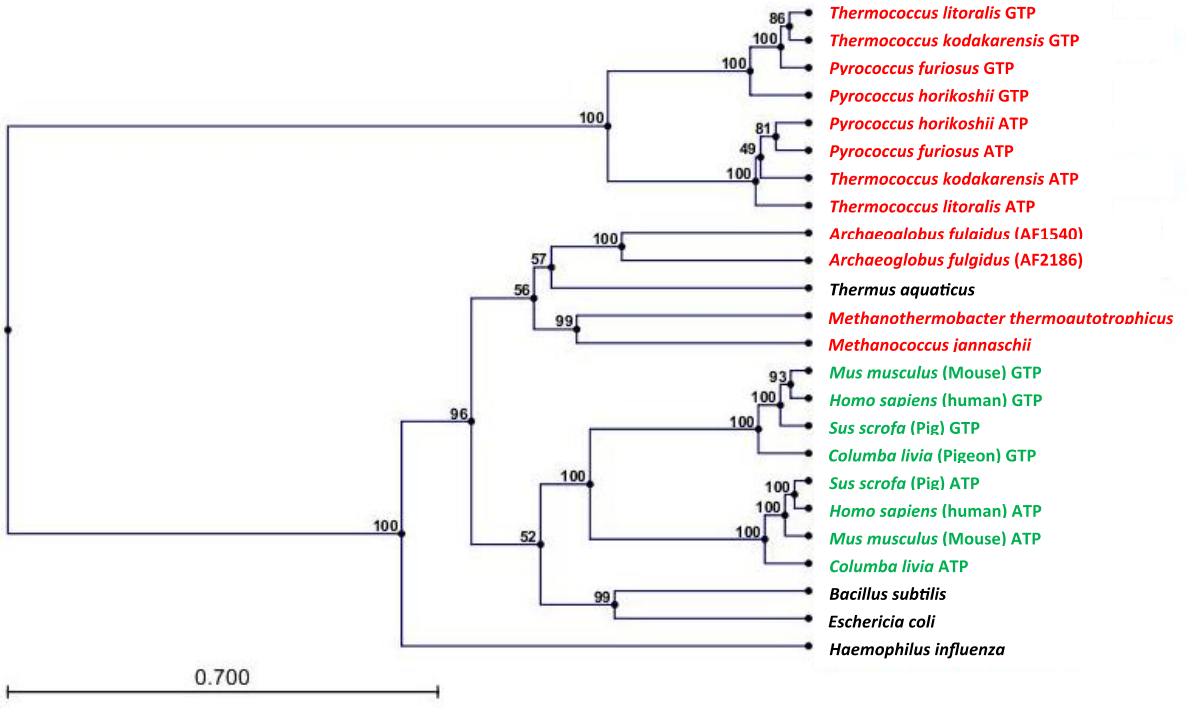
Homologs of ACS, known as succinyl-CoA synthetases (SCS), have been characterized from eukaryotes as well as bacteria [33; 34]. As with the *P. furiosus* enzymes, their nucleotide specificity is determined by their β subunits [35], although these differ from those in the ACSs of the Thermococcales as they contain an additional ligase domain [26; 36]. A phylogenetic tree of the β-subunits of the ACSs and SCSs is shown in Figure 3.4. It is striking that those of the Thermococcales form one clade while all other β-subunits fall into a second widely divergent clade. Each clade contains subunits that have a preference for one or the other purine nucleotide,

indicating that they each arose in each clade by independent gene duplications. Interestingly, the ACSs that are found in Archaea that are not members of the Thermococcales, such as those in some methanogens and in the hyperthermophilic sulfate-reducer *Archaeoglobus fulgidus*, fall within a distinct branch within the eukaryote/bacterial clade of the tree rather than in the Thermococcales branch (Figure 3.4). Clearly, the β-subunits found in the ACSs of the Thermococcales are very distinct from those in all other ACSs/SCSs found throughout biology, including those in other hyperthermophilic archaea. Why and how this occurred are beyond the scope of this paper but are interesting subjects worthy of additional study.

Figure 3.4.

Phylogenetic tree of ACS β subunits.

The unweighted pair group method of arithmetic mean (UPGMA) tree is shown. A neighbor-joining tree (data not shown) also shows similar relationships for the vertebrate and Thermococcales subunits with a few changes for the remaining species. Archaea are highlighted red, Bacteria are in black and Eukarya in green. Tree and Bootstrap values were generated using CLC Main Work Bench 6 with 100 resamples.


Thermococcus litoralis GTP
Thermococcus kodakarensis GTP
Pyrococcus furiosus GTP
Pyrococcus horikoshii GTP
Pyrococcus horikoshii ATP
Pyrococcus furiosus ATP
Thermococcus kodakarensis ATP
Thermococcus litoralis ATP
Archaeoglobus fulgidus (AF1540)
Archaeoglobus fulgidus (AF2186)
Thermus aquaticus
Methanothermobacter thermoautotrophicus
Methanococcus jannaschii
Mus musculus (Mouse) GTP
Homo sapiens (human) GTP
Sus scrofa (Pig) GTP
Columba livia (Pigeon) GTP
Sus scrofa (Pig) ATP
Homo sapiens (human) ATP
Mus musculus (Mouse) ATP
Columba livia ATP
Bacillus subtilis
Eschericia coli
Haemophilus influenza
0.700

References

1. Stetter, K.O., "Hyperthermophilic procaryotes," *FEMS Microbiol Rev*, vol. 18, no. 2-3, pp. 149-158, 1996.

2. de Vos, W.M., S.W.M. Kengen, W.G.B. Voorhorst and J. van der Oost, "Sugar utilization and its control in hyperthermophiles," *Extremophiles*, vol. 2, no. 3, pp. 201-205, 1998.

3. Verhees, C.H., S.W. Kengen, J.E. Tuininga, G.J. Schut, M.W. Adams, W.M. De Vos and J. Van Der Oost, "The unique features of glycolytic pathways in Archaea," *Biochem J*, vol. 375, pp. 231-246, 2003.

4. Kengen, S.W., F.A. de Bok, N.D. van Loo, C. Dijkema, A.J. Stams and W.M. de Vos, "Evidence for the operation of a novel Embden-Meyerhof pathway that involves ADP-dependent kinases during sugar fermentation by *Pyrococcus furiosus*," *J Biol Chem*, vol. 269, no. 26, pp. 17537-17541, 1994.

5. Mukund, S. and M.W. Adams, "Glyceraldehyde-3-phosphate ferredoxin oxidoreductase, a novel tungsten-containing enzyme with a potential glycolytic role in the hyperthermophilic archaeon Pyrococcus furiosus," *J Biol Chem*, vol. 270, no. 15, pp. 8389-92, 1995.

6. Siebers, B. and P. Schönheit, "Unusual pathways and enzymes of central carbohydrate metabolism in Archaea," *Curr Opin Microbiol*, vol. 8, no. 6, pp. 695-705, 2005.

7. Blamey, J.M. and M.W. Adams, "Purification and characterization of pyruvate ferredoxin oxidoreductase from the hyperthermophilic archaeon *Pyrococcus furiosus*," *Biochim Biophy Acta*, vol. 1161, no. 1, pp. 19-27, 1993.

8. Mai, X. and M.W. Adams, "Purification and characterization of two reversible and ADP-dependent acetyl coenzyme A synthetases from the hyperthermophilic archaeon *Pyrococcus furiosus*," *J Bacteriol*, vol. 178, no. 20, pp. 5897-5903, 1996.

9. Adams, M.W. and A. Kletzin, "Oxidoreductase-type enzymes and redox proteins involved in fermentative metabolisms of hyperthermophilic Archaea," *Adv Protein Chem*, vol. 48, pp. 101-180, 1996.

10. Mai, X. and M.W. Adams, "Indolepyruvate ferredoxin oxidoreductase from the hyperthermophilic archaeon *Pyrococcus furiosus*. A new enzyme involved in peptide fermentation," *J Biol Chem*, vol. 269, no. 24, pp. 16726-16732, 1994.

11. Heider, J., X. Mai and M.W. Adams, "Characterization of 2-ketoisovalerate ferredoxin oxidoreductase, a new and reversible coenzyme A dependent enzyme involved in peptide fermentation by hyperthermophilic archaea," *J Bacteriol*, vol. 178, no. 3, pp. 780-787, 1996.

12. Mai, X. and M.W. Adams, "Characterization of a fourth type of 2-keto acid-oxidizing enzyme from a hyperthermophilic archaeon: 2-ketoglutarate ferredoxin oxidoreductase from *Thermococcus litoralis*," *J Bacteriol*, vol. 178, no. 20, pp. 5890-5896, 1996.

13. Schut, G.J., S.D. Brehm, S. Datta and M.W. Adams, "Whole-genome DNA microarray analysis of a hyperthermophile and an archaeon: *Pyrococcus furiosus* grown on carbohydrates or peptides," *J Bacteriol*, vol. 185, no. 13, pp. 3935-3947, 2003.

14. Menon, A.L., F.L. Poole, 2nd, A. Cvetkovic, S.A. Trauger, E. Kalisiak, J.W. Scott, S. Shanmukh, J. Praissman, F.E. Jenney, Jr., W.R. Wikoff, J.V. Apon, G. Siuzdak and M.W. Adams, "Novel multiprotein complexes identified in the hyperthermophilic archaeon Pyrococcus furiosus by non-denaturing fractionation of the native proteome," *Mol Cell Proteomics*, vol. 8, no. 4, pp. 735-751, 2009.

15. Glasemacher, J., A.K. Bock, R. Schmid and P. Schønheit, "Purification and properties of acetyl-CoA synthetase (ADP-forming), an archaeal enzyme of acetate formation and ATP

synthesis, from the hyperthermophile *Pyrococcus furiosus*," *Eur J Biochem*, vol. 244, no. 2, pp. 561-567, 1997.

16. Musfeldt, M., M. Selig and P. Schönheit, "Acetyl coenzyme A synthetase (ADP forming) from the hyperthermophilic archaeon *Pyrococcus furiosus*: identification, cloning, separate expression of the encoding genes, acdAI and acdBI, in *Escherichia coli*, and in vitro reconstitution of the active heterotetrameric enzyme from its recombinant subunits," *J Bacteriol*, vol. 181, no. 18, pp. 5885-5888, 1999.

17. Robb, F.T., D.L. Maeder, J.R. Brown, J. DiRuggiero, M.D. Stump, R.K. Yeh, R.B. Weiss and D.M. Dunn, "Genomic sequence of hyperthermophile, *Pyrococcus furiosus*: implications for physiology and enzymology," *Methods Enzymol*, vol. 330, pp. 134-157, 2001.

18. Shikata, K., T. Fukui, H. Atomi and T. Imanaka, "A novel ADP-forming succinyl-CoA synthetase in *Thermococcus kodakaraensis* structurally related to the archaeal nucleoside diphosphate-forming acetyl-CoA synthetases," *J Biol Chem*, vol. 282, no. 37, pp. 26963-26970, 2007.

19. Awano, T., A. Wilming, H. Tomita, Y. Yokooji, T. Fukui, T. Imanaka and H. Atomi, "Characterization of Two Members among the Five ADP-Forming Acyl Coenzyme A (Acyl-CoA) Synthetases Reveals the Presence of a 2-(Imidazol-4-yl)Acetyl-CoA Synthetase in Thermococcus kodakarensis," *J Bacteriol*, vol. 196, no. 1, pp. 140-7, 2014.

20. Bradford, M.M., "Rapid and sensitive method for quantitation of microgram quantities of protein utilizing principle of protein-dye binding," *Anal Biochem*, vol. 72, no. 1-2, pp. 248-254, 1976.

21. Raven, N. and R.J. Sharp, "Development of defined and minimal media for the growth of the hyperthermophilic archaeon *Pyrococcus furiosus Vc1*.," *FEMS Microbiol Lett*, vol. 146, no. 1, pp. 135-141, 1997.

22. Blumentals, II, S.H. Brown, R.N. Schicho, A.K. Skaja, H.R. Costantino and R.M. Kelly, "The hyperthermophilic archaebacterium, Pyrococcus furiosus. Development of culturing protocols, perspectives on scaleup, and potential applications," *Ann NY Acad Sci*, vol. 589, pp. 301-14, 1990.

23. Hoaki, T., M. Nishijima, M. Kato, K. Adachi, S. Mizobuchi, N. Hanzawa and T. Maruyama, "Growth requirements of hyperthermophilic sulfur-dependent heterotrophic archaea isolated from a shallow submarine geothermal system with reference to their essential amino acids," *App Environ Microb*, vol. 60, no. 8, pp. 2898-904, 1994.

24. Ward, D.E., S.W. Kengen, J. van Der Oost and W.M. de Vos, "Purification and characterization of the alanine aminotransferase from the hyperthermophilic Archaeon pyrococcus furiosus and its role in alanine production," *J Bacteriol*, vol. 182, no. 9, pp. 2559-66, 2000.

25. Musfeldt, M. and P. Schönheit, "Novel type of ADP-forming acetyl coenzyme A synthetase in hyperthermophilic archaea: heterologous expression and characterization of isoenzymes from the sulfate reducer *Archaeoglobus fulgidus* and the methanogen *Methanococcus jannaschii*," *J Bacteriol*, vol. 184, no. 3, pp. 636-644, 2002.

26. Bräsen, C., M. Schmidt, J. Grötzinger and P. Schönheit, "Reaction mechanism and structural model of ADP-forming Acetyl-CoA synthetase from the hyperthermophilic archaeon *Pyrococcus furiosus*: evidence for a second active site histidine residue," *J Biol Chem*, vol. 283, no. 22, pp. 15409-15418, 2008.

27. Sapra, R., K. Bagramyan and M.W. Adams, "A simple energy-conserving system: proton reduction coupled to proton translocation," *P Natl Acad Sci USA*, vol. 100, no. 13, pp. 7545-7550, 2003.

28. Verhaart, M.R., A.A. Bielen, J. van der Oost, A.J. Stams and S.W. Kengen, "Hydrogen production by hyperthermophilic and extremely thermophilic bacteria and archaea: mechanisms for reductant disposal," *Environ Technol*, vol. 31, no. 8-9, pp. 993-1003, 2010.

29. Mukund, S. and M.W. Adams, "Glyceraldehyde-3-phosphate ferredoxin oxidoreductase, a novel tungsten-containing enzyme with a potential glycolytic role in the hyperthermophilic archaeon *Pyrococcus furiosus*," *J Biol Chem*, vol. 270, no. 15, pp. 8389-8392, 1995.

30. Fukuda, W., T. Fukui, H. Atomi and T. Imanaka, "First characterization of an archaeal GTP-dependent phosphoenolpyruvate carboxykinase from the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1," *J Bacteriol*, vol. 186, no. 14, pp. 4620-4627, 2004.

31. Matte, A., L.W. Tari, H. Goldie and L.T. Delbaere, "Structure and mechanism of phosphoenolpyruvate carboxykinase," *J Biol Chem*, vol. 272, no. 13, pp. 8105-8108, 1997.

32. Garrett, R.A. and H. Klenk. 2007. Translational mechanisms and protein synthesis. *In* Archaea: evolution, physiology, and molecular biology. Blackwell Publishing, Malden, MA, pp 217-230.

33. Lambeth, D.O., K.N. Tews, S. Adkins, D. Frohlich and B.I. Milavetz, "Expression of two succinyl-CoA synthetases with different nucleotide specificities in mammalian tissues," *J Biol Chem*, vol. 279, no. 35, pp. 36621-36624, 2004.

34. Johnson, J.D., W.W. Muhonen and D.O. Lambeth, "Characterization of the ATP- and GTP-specific succinyl-CoA synthetases in pigeon. The enzymes incorporate the same alpha-subunit," *J Biol Chem*, vol. 273, no. 42, pp. 27573-27579, 1998.

35. Johnson, J.D., J.G. Mehus, K. Tews, B.I. Milavetz and D.O. Lambeth, "Genetic evidence for the expression of ATP- and GTP-specific succinyl-CoA synthetases in multicellular eucaryotes," *J Biol Chem*, vol. 273, no. 42, pp. 27580-27586, 1998.

36. Sànchez, L.B., M.Y. Galperin and M. Müller, "Acetyl-CoA synthetase from the amitochondriate eukaryote *Giardia lamblia* belongs to the newly recognized superfamily of acyl-CoA synthetases (Nucleoside diphosphate-forming)," *J Biol Chem*, vol. 275, no. 8, pp. 5794-5803, 2000.

CHAPTER 4

IDENTIFYING NOVEL METALLOPROTEINS IN THE PROTEOME OF THE HYPERTHERMOPHILIC ARCHAEON *PYROCOCCUS FURIOSUS*

Introduction

Metalloproteins are very significant in biology. Some estimates suggest that metals are an integral part of up to 40 % of all proteins [1]. The incorporation of metals into proteins contributes to protein stability and confers enzymes with virtually unlimited catalytic potential. Examples of major processes in which metalloproteins participate are, respiration (e.g. cytochrome oxidase [Fe/Cu]), photosynthesis (e.g. photosystem II [Mn]), and nitrogen assimilation (e.g. nitrate reductase [Mo], and nitrogenase [Fe/Mo]). The knowledge that a protein contains a metal ion or metal associated prosthetic group greatly advances the understanding of the protein's structure and function. Consequently, knowing what metals bind to what proteins advances our knowledge of biochemical pathways, our understanding of human disease and is essential to progress in bioremediation efforts. Hence, metalloproteomes are great stores of vital information. However, only a limited number of proteins that have been characterized are metalloproteins. Due to the diversity of metalloproteins, the nature of the sites that bind metals, and the small sample size of characterized metalloproteins, it is not often presently possible to meaningfully predict the number and types of metalloproteins in any organism from its genome sequence. Even if a putative metal binding site is determined, it is not obvious which metal

binds to it. For example, both zinc and iron sulfur clusters are bound by motifs containing multiple cysteinyl residues that may not be easily differentiated [2].

Historically the identification of metalloproteins has tended to be of secondary concern, and the presence of a metal ion in a particular protein may not even be suspected until the protein is purified and subsequently analyzed. Given that only a small fraction of proteins have been characterized biochemically and that the metal contents of proteins are not routinely analyzed, coupled with the limitations of metalloprotein informatics, the true extent of the metalloproteome of any organism is clearly under defined. Some studies have focused on the metal contents of either purified proteins or of biological material such as blood or urine in response to environmental impact or disease [3], rather than on genome-wide approaches. Other studies have used high throughput colorimetric methods for metalloproteome analysis but involved a limited number of samples [4]. Still other studies included analyses of recombinant biological samples such as in structural genomics initiatives [5]. Analyses of recombinant proteins produced in structural genomics are prone to misinterpretation as the host organism used to produce the recombinant protein by heterologous expression may insert the wrong metal [6] or no metal at all. It is therefore essential that native biomass be analyzed for potential metalloproteins. Thus far, a genome wide analysis of the native metalloproteome of an organism has yet to be performed. Therefore, the native metalloproteome of any organism has never been characterized.

Herein we investigated the metalloproteome of the model organism *P. furiosus*. Thirty-four metalloproteins have thus far been purified from *P. furiosus* native biomass with most containing iron (Fe), while others contain nickel (Ni), tungsten (W), cobalt (Co) and zinc (Zn) Table 4.1. For the most part, these proteins were purified in the typical way whereby, the

Table 4.1. Characterized metalloproteins of *P. furiosus*.

No.	Gene name	Protein annotation	Protein expression form	Metal	Reference
1	PF0080	tungsten transport protein A WtpA	recombinant	W	[8]
2	PF0094	disulfide oxidoreductases	recombinant	Zn	[9]
3	PF0196	glucose-6-phosphate isomerase	native	Fe	[10]
4	PF0346	aldehyde:ferredoxin oxidoreductase	native	W, Fe	[11]
5	PF0440	ribonucleotide reductase	native	Co	[12]
6	PF0456	cobalt-activated carboxypeptidase	native	Co	[13]
7	PF0464	glyceraldehyde-3-phosphate ferredoxin oxidoreductase	native	W, Fe	[14]
8	PF0477	alpha-amylase	recombinant	Zn	[15]
9	PF0533 PF0534	indolepyruvate ferredoxin oxidoreductase	native	Fe	[16]
10	PF0541	methionine aminopeptidase	recombinant	Co, Mn, Fe	[17-19]
11	PF0549	hydrogenase maturation protein HypD	recombinant	Fe	[20]
12	PF0597	Aminoacylase	native	Co, Zn	[21]
13	PF0610	winged helix-turn-helix variant	recombinant	Zn, Fe	[22]
14	PF0742	Ferritin	native	Fe, Zn	[23]
15	PF0891 PF0892 PF0893 PF0894	hydrogenase I	native	Ni,Fe	[24]
16	PF0965 PF0966 PF0967 PF0971	pyruvate ferredoxin oxidoreductase	native	Fe	[25]
17	PF0968 PF0969 PF0970 PF0971	2-ketoisovalerate ferredoxin oxidoreductase	native	Fe	[26]
18	PF0991	threonine dehydrogenase	recombinant	Zn	[27]

Table 4.1. continued.

19	PF1167	Rad50	recombinant	Zn	[28]
20	PF1193	Dps-like protein	recombinant	Fe	[29]
21	PF1203	formaldehyde:ferredoxin oxidoreductase	Native	W, Fe	[30]
22	PF1281	superoxide reductase	Native	Fe	[31]
23	PF1282	Rubredoxin	Native	Fe	[32]
24	PF1283	Rubrerythrin	Native	Fe	[33]
25	PF1327 PF1328	ferredoxin NADPH oxidoreductase (FNOR I)	Native	Fe	[34]
26	PF1329 PF1330 PF1331 PF1332	hydrogenase II	Native	Fe	[35]
27	PF1343	proline dipeptidase	Native	Co	[36]
28	PF1433 PF1434	mbh12 hydrogenase	Native	Ni, Fe	[37]
29	PF1480	formaldehyde:ferredoxin oxidoreductase wor5	Native	W, Fe	[38]
30	PF1861	lysyl aminopeptidase	Native	Zn, Co	[39]
31	PF1909	Ferredoxin	Native	Fe	[40]
32	PF1910 PF1911	sulfide dehydrogenase	Native	Fe	[41]
33	PF1953	N-terminal cystathionine beta synthase	recombinant	Fe (Zn)	[42]
34	PF1961	tungsten-containing formaldehyde ferredoxin oxidoreductase wor4	Native	W, Fe	[43]

presence of a metal was determined after purification by following enzyme activity. One exception is aldehyde ferredoxin oxidoreductatse (PF0346) once deemed RTP (red tungsten protein) [7]. This enzyme was first purified by following a red color during column chromatography due to the presence of iron and tungsten chromophores in the enzyme. Here we present an approach that is also independent of enzyme activity, however, this method detects, identifies and purifies metalloproteins on a genome-wide scale. It is based on the native purification of metalloproteins and detects them by following metals through a series of chromatographic steps while using protein mass spectrometry to identify the proteins that co-elute with the metals. The metals are assayed by inductively coupled mass spectrometry (ICP-MS). ICP-MS is highly sensitive, has high isotope specificity, has a large dynamic range, detects almost all elements [44; 45], and because it can perform simultaneous multi-elemental analysis is conducive to high throughput analysis. Furthermore, because this method does not discriminate on the basis of whether an organism can be genetically manipulated, it can be applied to any biological system. Like the method used to identify novel protein-protein interactions [46], the method described here also uses peptide mass spectrometry to assign proteins to the metal peaks delineated by ICP-MS and thus predicts which proteins bind which metals.

The power of this method is demonstrated in the detection of possibly hundreds of uncharacterized metalloproteins in *P. furiosus.* Many of the uncharacterized metalloproteins had metals already known to be present in *P. furiosus* proteins, such as Fe, Ni, W, Co, and Zn. However, there is also evidence for metalloproteins containing (V) vanadium, (Mo) molybdenum, (Mn) manganese, and Pb (lead) that have not previously been known to be present in *P. furiosus* proteins. For example, the only nickel enzymes described in *P. furiosus* are hydrogenases [37; 47], however multiple potential nickel binding proteins were detected in its

biomass which could not be accounted for. The purification of one of these novel nickel proteins and a novel molybdenum binding protein will be the subject of the following chapter of this thesis. However, this chapter focuses on the application of this method with respect to a global view of the metalloproteome of *P. furiosus*.

Materials and Methods.

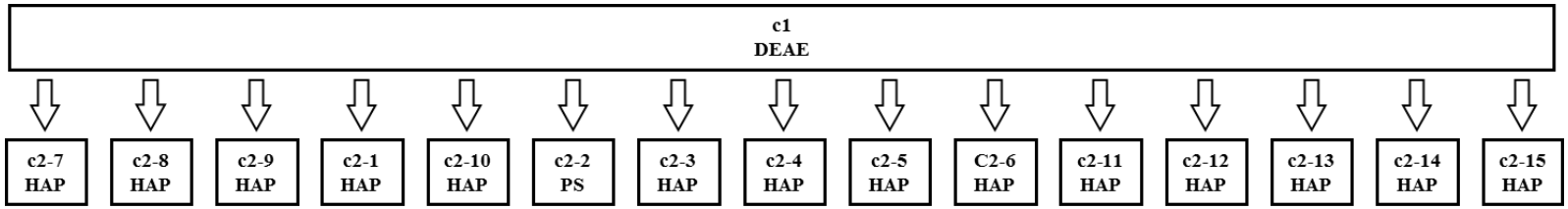
Preparation of anaerobic cell-free extract**.** *P. furiosus* (DSM 3638) was grown under anaerobic, reducing conditions at 90°C in a 600-liter fermenter on maltose and peptides and harvested in late log phase [46]. The procedures for preparing anaerobic cell-free extract and for running the first chromatography step have been described previously [46]. Briefly, 300 g of frozen cells were gently lysed by osmotic shock anaerobically under a continuous flow of argon in 900 ml of 50mM Tris-HCl (pH 8.0) containing 2mM sodium dithionite as a reductant (Buffer A) and 0.5 μg/ml DNAse I to reduce viscosity. The cell lysate was centrifuged at 100,000 x g for 2 h at 18°C, and the clear supernatant, representing the cytoplasmic fraction was immediately loaded onto the first column. This and all subsequent columns were run under anaerobic, reducing conditions where all buffers were degassed and maintained under a positive pressure of argon, and all liquid transfers were made using needles and syringes.

Large scale anaerobic fractionation of cytoplasmic proteins. *First column (c1) fractionation.* A column tree representing the complete fractionation is shown in Figure 4.1. All columns were run using an ÄKTA^{TM} basic automated LC system (GE Healthcare). The *P. furiosus* cytoplasmic fraction was diluted 4-fold in Buffer A to reduce the ionic strength and was loaded onto a 10 X 20-cm (1.5-liter) column of DEAE-Sepharose Fast Flow (GE Healthcare) that had been equilibrated previously with the same buffer until the effluent was anaerobic and reducing. Unbound protein was washed off the column with Buffer A. Bound proteins were eluted using a

Figure 4.1

Fractionation scheme of *P. furiosus* cytoplasmic proteins through primary and secondary purification steps.

Cytoplasmic proteins were separated on a DEAE anion exchange column which was the first fractionation. Fractions from the DEAE column were combined into 15 pools that were fractionated through subsequent columns. The chromatography steps included DEAE (diethylaminoethyl cellulose), HAP (hydroxyapatite), PS (phenyl sepharose), BS (blue sepharose), and SX200 (superdex 200) which are indicated in the figure. Details of secondary fractionations are given in reference 46. The fractions generated were analyzed by either CV or HT or both. Fractions were analyzed by both ICP-MS and peptide mass spectrometry.


c1
DEAE
c2-7
HAP
c2-8
HAP
c2-9
HAP
c2-1
HAP
c2-10
HAP
c2-2
PS
c2-3
HAP
c2-4
HAP
c2-5
HAP
C2-6
HAP
c2-11
HAP
c2-12
HAP
c2-13
HAP
c2-14
HAP
c2-15
HAP

linear gradient (15 liters) of 0-500 mM NaCl in Buffer A, and 125-ml anaerobic fractions were collected (120 fractions). Any remaining tightly bound proteins were eluted from the column using 1-liter step gradients of 1, 1.5, and 2 M NaCl respectively, in Buffer A (high salt washes).

Secondary (c2) chromatographic steps. A subset of 37 gradient fractions from the first column separation were selected using a combination of protein profiles and enzyme activities to create 6 fraction pools, c1-1 through c1-6. The control proteins and enzyme activities were as follows: c1-1, ferredoxin-NADPH oxidoreductase (FNOR) [34] and rubrerythrin [33]; c1-2, 2-ketisovalerate ferredoxin oxidoreductase (VOR) and pyruvate ferredoxin reductase (POR) [48]; c1-3, aldehyde:ferredoxin oxidoreductase [49]; c1-4, soluble hydrogenase 1 (SH1) [24]; c1-5, rubredoxin [32]; and c1-6 ferredoxin (Fd)[40]. These pools, c1-1 through c1-6, were fractionated through six subsequent c2 columns. The remaining 83 gradient fractions were combined to generate nine additional pools, in this case based solely on SDS-PAGE and protein profiles. These pools c1-7 through c1-15, were fractionated through nine subsequent c2 columns to give a total of 15 columns at the c2 level (Figure 4.1). SDS-PAGE and native PAGE were carried out using precast 4-20 % gradient gels (Criterion gel system, Bio-Rad). Protein concentrations were estimated by the method of Bradford [50] using bovine serum albumin as the standard.

Protein identification. The procedures for protein identification in solution using high-throughput tandem mass spectrometry (HT-MS/MS) were described previously [46]. Mass list generation and data base searches are described in reference 51.

Metal analyses. Metal analysis was done by Aleks Cvetokic and Saratchandra Shanmukh using ICP-MS. A detailed description of the method in ref 51. Identification of metalloproteins in *P. furiosus***.** Farris Poole, William Lancaster, and Jeremy Praissman carried out metalloprotein

bioinformatics analysis. A detailed description of the method is given [51].

Fractionation of *S solfataricus* and *E. coli*. Angeli Menon of the University of Georgia and a collaborator, Dr.Steve Yanonne, of Lawrence Berkeley National Laboratory grew and fractionated *S. solfataricus*. Angeli Menon and Michael Thorgersen of the University of Georgia grew and fractionated *E. coli*. A detailed description of the growth is in the reference [51].

Results and Discussion

Analysis of *P. furiosus* cell free extract indicates that twenty-one elements of the forty-four found in the growth medium are assimilated by *P. furiosus* into its cytoplasmic fraction. Ultrafiltration experiments using a pore size of 5kDa indicated that eighteen of the metals bind to proteins while (Cr) chromium, (Sr) strontium, and (Ru) ruthenium do not. In order to identify proteins that tightly bind metals, the cytoplasmic extract was loaded onto an anion exchange column and fractions were collected and analyzed for metal content along with protein content. Of the 54 elements assayed in the fractions by ICP-MS, distinct peaks were detected in the fractions for 10 metals (Co, Fe, Mn, Mo, Ni, Pb, U, V, W, and Zn). These metals remained bound to the column through extensive washing in buffer of low ionic strength and eluted only after the ionic strength was increased. It is assumed that the metals are bound to *P. furiosus* proteins either covalently or by strong ionic interactions. Co, Fe, Ni, W, and Zn-proteins have been previously identified in *P. furiosus*, however, Mo, Mn, and V containing proteins have not, and with the exception of detoxification proteins, Pb and U proteins have not been identified in any organism. Interestingly, V, Pb, and Mo were not specifically added to the growth medium [46], but had peaks detected for each metal and were at concentrations of 0.602 ± 0.0209 μM, 0.031 ± 0.001 μM, and 0.118 ± 0.006 μM, respectively. It is assumed these metals are impurities that were derived from organic components (e.g. tryptone and yeast extract) in the media. The

observation of uptake and assimilation of these metals suggests that they have biological roles in *P. furiosus.*

The concentration profiles of V, Pb, Co, Ni, Mn, Mo, U, and W are shown in Figure 4.2 (Fe and Zn are not shown). Peptide mass spectrometry of the DEAE fractions identified 696 distinct proteins. A total of 76 metal peaks were also detected in those fractions, but twenty-three of the metal peaks could not be assigned to known metalloproteins. The DEAE fractions were combined into 15 separate pools and each was subjected to further purification in order to further analyze the metal peaks. Secondary purifications resulted in 790 new fractions. Peptide mass spectrometry identified 770 proteins in these fractions. The secondary purifications resolved the 76 metal peaks of the DEAE fractionation into 343 metal peaks. The total number of peaks for each metal from the primary and secondary columns is summarized in Table 4.2. Of these, only 185 could be assigned to known metalloproteins or proteins that have a predicted metal binding domain. The remaining 158 peaks with no assignment thus, represent novel metalloproteins having uncharacterized metal binding domains. Eight of the unassigned metal peaks were further purified in an attempt to determine what protein they were bound to. After three chromatography steps PF0742, annotated as ferritin, was purified as a homogeneous fraction with a 1.20 ± 0.11 Fe atoms mol^{-1} protein, but with only 0.010 ± 0.001 U atoms mol^{-1} protein. PF0215, annotated as an enolase, was purified through six steps with another uranium peak. The metal stoichiometry for this protein was 0.00010 ± 0.00004 U atoms mol^{-1} protein. PF1343, a known metalloprotease contained 0.65 ± 0.09 Zn atoms mol^{-1}, however it contained substoichiometric amounts of lead (0.01 ± 0.002 Pb atoms mol^{-1} protein). Similarly, purification

Figure 4.2.

Metal concentration profiles of *P. furiosus* extract after DEAE-sepharose fractionation of cytoplasmic extract.

Cytoplasmic extract was separated on a DEAE anion exchange column. The fractions were analyzed by ICP-MS to determine the concentrations of (a) vanadium (V) and lead (Pb); (b) nickel (Ni) and cobalt (Co); (c) manganese (Mn) and molybdenum (Mo); and (d) uranium (U) and tungsten (W). The arrows indicate metal peaks detected.

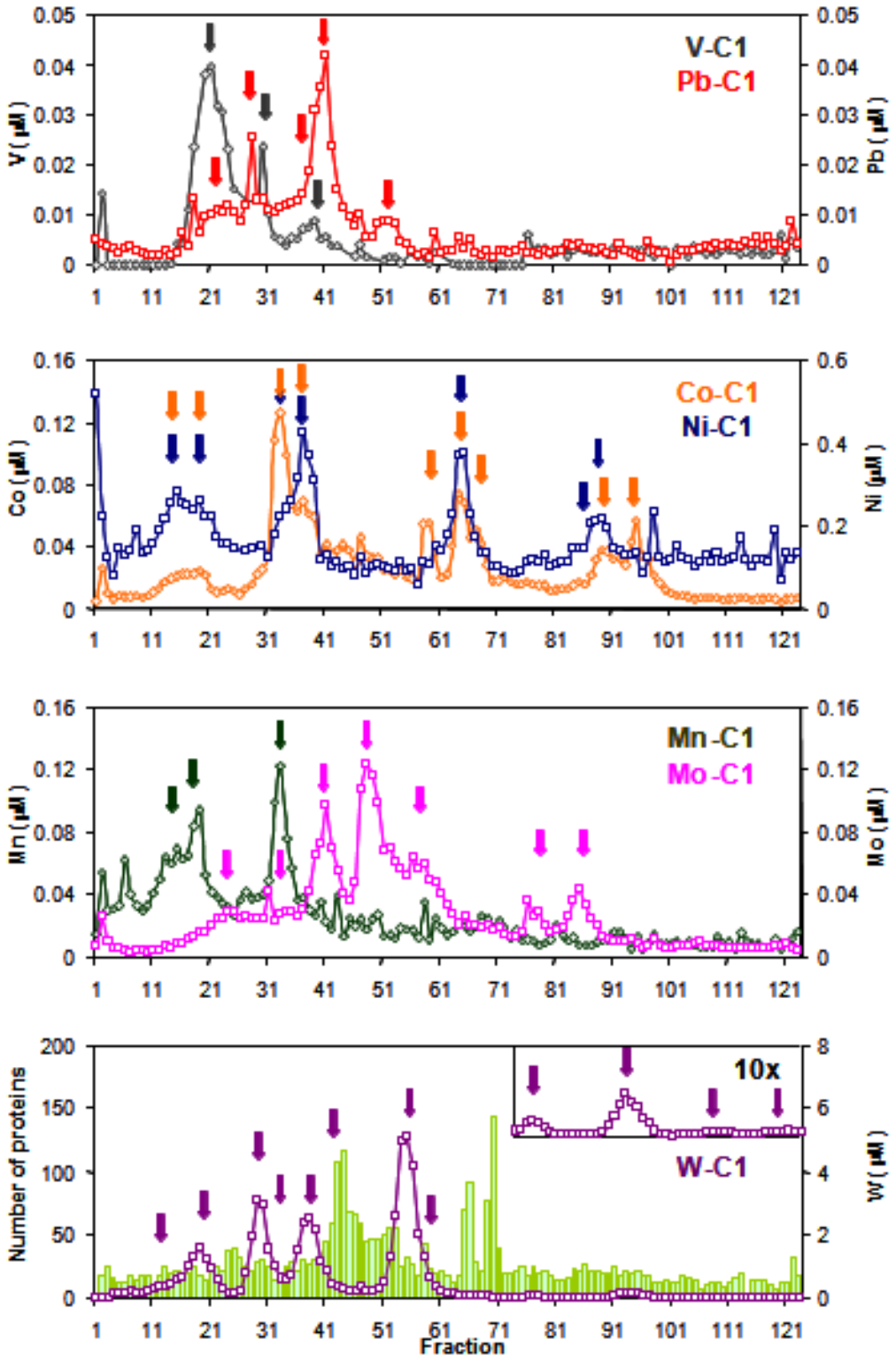

V-C1
Pb-C1
V (μM)
Pb (μM)
Co-C1
Ni-C1
Co (μM)
Ni (μM)
Mn-C1
Mo-C1
Mn (μM)
Mo (μM)
10x
W-C1
W (μM)
Number of proteins
Fraction

Table 4.2. Number of metal peaks detected from primary and secondary fractionations. The number of peaks for a given metal is indicated for each column. No value indicates the absence of metal peak. Numbers in parentheses indicates peaks for which a protein can be assigned. Uranium peaks were not purified past the first column step. Adapted from (ref)

Metal	C1 totals	C2-1 HAP	C2-2 PS	C2-3 HAP	C2-4 HAP	C2-5 HAP	C2-6 HAP	C2-7 HAP	C2-8 HAP	C2-9 HAP	C2-10 HAP	C2-11 HAP	C2-12 HAP	C2-13 HAP	C2-14 HAP	C2-15 HAP	C2 totals
Fe	14(14)	7(7)	2(1)	5(5)	7(7)	6(6)	4(4)		5(5)	4(4)	3(3)	6(6)	3(3)	1(1)	4(3)	1(1)	58(56)
Co	13(11)	7(5)	6(1)	3(2)	5(2)	8(3)	2(1)	3(0)	6(0)	3(1)	3(3)	3(1)	4(3)	3(0)	4(3)	2(1)	62(26)
W	13(12)	5(5)	3(1)	5(5)	4(3)	4(2)	4(2)	2(1)	7(6)	4(4)	2(2)	5(4)	5(2)		1(1)	1(1)	52(39)
Ni	7(2)	5(0)	1(0)	5(1)	3(2)	5(4)		2(0)	2(1)	4(0)	4(0)	3(1)	4(4)	3(0)		1(0)	42(13)
Mo	8(6)	5(5)	2(1)	4(0)	6(0)			1(0)	2(2)	2(1)	5(3)	2(1)	4(2)		2(1)	2(2)	37(19)
Zn	6(6)			4(4)	4(4)			4(2)		2(2)	2(2)	3(3)	4(4)	4(3)	1(1)		28(25)
Pb	5(0)	7(0)		7(0)	1(0)	1(0)			4(0)	1(0)	2(0)	2(0)	6(0)		1(0)	2(0)	34(0)
Mn	4(2)	4(2)	2(0)					2(0)	2(1)	3(2)		2(1)	4(0)		2(1)		21(7)
V	3(0)								3(0)	2(0)	1(0)						6(0)
U	3(0)																3(0)
Total metal peaks in column	76(53)	40(24)	16(4)	33(17)	30(18)	24(16)	10(7)	14(3)	31(15)	25(14)	22(13)	26(17)	34(18)	11(4)	15(10)	9(5)	343(185)

of an additional Pb peak through five columns yielded PF0257, a pyrophosphatase. While this protein had an Fe stoichiometry of 1.40 ± 0.02, its Pb stoichiometry was relatively low at 0.007 ± 0.001. Thus, PF0257 is apparently a new iron protein. Nickel and molybdenum binding proteins were also purified and identified. Unlike the lead and uranium proteins these contained higher amounts of metals. PF0056 is a cupin/putative sugar-binding protein that tightly binds nickel through five chromatography steps. Although it has no predicted nickel binding domain, it contained 0.47 ± 0.05 Ni atoms mol^{-1} protein and 0.50 ± 0.08 Zn atoms mol^{-1}. The nickel atom is predicted to be coordinated by three His residues and one Glu residue as suggested from a structure of a homologous protein. PF1587 purified with a molybdenum and known tungstoprotein (PF0464) through 6 chromatography steps. PF0464 was found to lack molybdenum, therefore it was concluded that PF1587 was the protein that binds molybdenum. The discovery of new nickel and molybdenum binding proteins is significant because in *P. furiosus*, nickel enzymes have thus far been only known to be involved in hydrogen metabolism and in *P. furiosus* tungsten as a rule is used rather than molybdenum in enzymes. An additional nickel binding protein (PF0086) and molybdenum binding protein (PF1972) were extensively examined and the results of that study are the subject of chapter 5.

The method used herein to characterize metalloproteomes was further validated in experiments with extracts of *Escherichia coli* and *Sulfolobus solfataricus*. *E. coli* is a facultative anaerobic mesophile. Its environment is very different from that of *P. furiosus* in that it is common in the human gut rather than marine vents whereas, *S. solfataricus* thrives in a fresh water environment that is both aerobic and acidic. *E. coli* and *S. solfataricus* biomass was produced and the cell free extracts of each was fractionated off a DEAE-Sepharose column. Although, the growth media of *P. furiosus, E. coli,* and *S. solfataricus* all include the same 44

metals, fractionation revealed unique concentration profiles for each extract. Table 4.3 makes a tally of the metal peaks detected in each of the three microorganisms. Peaks for Co, Fe, Mo, Mn, V, Zn, and Pb were found in in the extracts of all species, but differed in their elution with respect to ionic strength. In addition, while *E. coli* uniquely demonstrated the presence of cadmium (Cd) and arsenic (As), *S. solfataricus* fractionation uniquely revealed tin (Sn) and antimony (Sb), barium (Ba), and thallium (Tl) peaks. Furthermore, like *P. furiosus, E. coli* fractions contained U and Ni peaks while those of *S. solfataricus* did not. *P. furiosus* was the only organism that had tungsten (W) peaks. Differing metal concentration profiles not only suggests that proteins in the three species interact differently with the metals present in the media, but that they also have differing preferences as to which metals they assimilate. This observation may reflect an adaptation to metal availability in their respective natural environments. Thus, an additional benefit of this method is that it may aid in the development of culture conditions that may be needed to grow fastidious microbes. In other words, it may be that some cultures do not grow in the lab because their proteins cannot function due to a lack of proper metals.

Improper metal insertion may affect a function of a protein in two ways. 1) It may disrupt its reaction mechanism or 2) it may render the protein structure unstable. Therefore, knowing what metals interact with a protein in its native form may prove beneficial in recombinant expression of proteins since the heterologous host may insert the wrong metal or no metal at all. Since uranium and lead have not been shown to have any metabolic function, their assimilation into cells and apparent capacity to bind to proteins may indicate that higher concentrations of these metals could contribute to metal toxicity.

Table 4.3 Metal peak detection in *P. furiosus, E. coli, and S. solfataricus.* ✓ indicate that a peak was detected and ✗ indicates a peak was not detected.

Metal	*P. furiosus*	*E. coli*	*S. solfataricus*
V	✓	✓	✓
Mn	✓	✓	✓
Fe	✓	✓	✓
Co	✓	✓	✓
Ni	✓	✓	✗
Cu	✗	✓	✓
Zn	✓	✓	✓
As	✗	✓	✗
Mo	✓	✓	✓
Cd	✗	✓	✗
Sn	✗	✗	✓
Ba	✗	✗	✓
W	✓	✗	✗
Tl	✗	✗	✓
Pb	✓	✓	✓
U	✓	✓	✗
Sb	✗	✗	✓

The approach described herein is generally applicable to a survey of the metalloproteome of any cell type. A comparison of the metal concentration profiles *P. furiosus*, *S. solfataricus* and *E. coli*, illustrates the differences in the metalloproteomes of these organisms. The use of peptide mass spectrometry along with column chromatrography, and ICP-MS, was used to assign proteins to metal peaks in *P. furiosus*. The fact that about half of the peaks could not be assigned to a protein demonstrates that metalloproteomes of organisms are poorly characterized.

References

1. Seravalli, J. and S.W. Ragsdale, "Expanding the biological periodic table," *Chem Biol*, vol. 17, no. 8, pp. 793-4, 2010.

2. Passerini, A., C. Andreini, S. Menchetti, A. Rosato and P. Frasconi, "Predicting zinc binding at the proteome level," *BMC Bioinformatics*, vol. 8, pp. 39, 2007.

3. Lobinski, R., D. Schaumloffel and J. Szpunar, "Mass spectrometry in bioinorganic analytical chemistry," *Mass Spectrom Rev*, vol. 25, no. 2, pp. 255-89, 2006.

4. Atanassova, A., M. Hogbom and D.B. Zamble, "High throughput methods for analyzing transition metals in proteins on a microgram scale," *Methods Mol Biol*, vol. 426, pp. 319-30, 2008.

5. Shi, W., C. Zhan, A. Ignatov, B.A. Manjasetty, N. Marinkovic, M. Sullivan, R. Huang and M.R. Chance, "Metalloproteomics: high-throughput structural and functional annotation of proteins in structural genomics," *Structure*, vol. 13, no. 10, pp. 1473-86, 2005.

6. Eidsness, M.K., S.E. O'Dell, D.M. Kurtz, Jr., R.L. Robson and R.A. Scott, "Expression of a synthetic gene coding for the amino acid sequence of Clostridium pasteurianum rubredoxin," *Protein Eng*, vol. 5, no. 4, pp. 367-71, 1992.

7. Mukund, S. and M.W.W. Adams, "Characterization of a Tungsten-Iron-Sulfur Protein Exhibiting Novel Spectroscopic and Redox Properties from the Hyperthermophilic Archaebacterium Pyrococcus-Furiosus," *Journal of Biological Chemistry*, vol. 265, no. 20, pp. 11508-11516, 1990.

8. Bevers, L.E., P.L. Hagedoorn, G.C. Krijger and W.R. Hagen, "Tungsten transport protein A (WtpA) in Pyrococcus furiosus: the first member of a new class of tungstate and molybdate transporters," *J Bacteriol*, vol. 188, no. 18, pp. 6498-505, 2006.

9. Ren, B., G. Tibbelin, D. de Pascale, M. Rossi, S. Bartolucci and R. Ladenstein, "A protein disulfide oxidoreductase from the archaeon Pyrococcus furiosus contains two thioredoxin fold units," *Nat Struct Biol*, vol. 5, no. 7, pp. 602-11, 1998.

10. Hansen, T., M. Oehlmann and P. Schonheit, "Novel type of glucose-6-phosphate isomerase in the hyperthermophilic archaeon Pyrococcus furiosus," *J Bacteriol*, vol. 183, no. 11, pp. 3428-35, 2001.

11. Mukund, S. and M.W. Adams, "Glyceraldehyde-3-phosphate ferredoxin oxidoreductase, a novel tungsten-containing enzyme with a potential glycolytic role in the hyperthermophilic archaeon *Pyrococcus furiosus*," *Journal of Biological Chemistry*, vol. 270, no. 15, pp. 8389-8392, 1995.

12. Riera, J., F.T. Robb, R. Weiss and M. Fontecave, "Ribonucleotide reductase in the archaeon Pyrococcus furiosus: a critical enzyme in the evolution of DNA genomes?," *Proc Natl Acad Sci U S A*, vol. 94, no. 2, pp. 475-8, 1997.

13. Cheng, T.C., V. Ramakrishnan and S.I. Chan, "Purification and characterization of a cobalt-activated carboxypeptidase from the hyperthermophilic archaeon Pyrococcus furiosus," *Protein Sci*, vol. 8, no. 11, pp. 2474-86, 1999.

14. Mukund, S. and M.W. Adams, "Glyceraldehyde-3-phosphate ferredoxin oxidoreductase, a novel tungsten-containing enzyme with a potential glycolytic role in the hyperthermophilic archaeon Pyrococcus furiosus," *J Biol Chem*, vol. 270, no. 15, pp. 8389-92, 1995.

15. Savchenko, A., C. Vieille, S. Kang and J.G. Zeikus, "Pyrococcus furiosus alpha-amylase is stabilized by calcium and zinc," *Biochemistry*, vol. 41, no. 19, pp. 6193-201, 2002.

16. Mai, X. and M.W. Adams, "Indolepyruvate ferredoxin oxidoreductase from the hyperthermophilic archaeon *Pyrococcus furiosus*. A new enzyme involved in peptide fermentation," *Journal of Biological Chemistry*, vol. 269, no. 24, pp. 16726-16732, 1994.

17. Copik, A.J., S. Waterson, S.I. Swierczek, B. Bennett and R.C. Holz, "Both nucleophile and substrate bind to the catalytic Fe(II)-center in the type-II methionyl aminopeptidase from Pyrococcus furiosus," *Inorg Chem*, vol. 44, no. 5, pp. 1160-2, 2005.

18. Tahirov, T.H., H. Oki, T. Tsukihara, K. Ogasahara, K. Yutani, K. Ogata, Y. Izu, S. Tsunasawa and I. Kato, "Crystal structure of methionine aminopeptidase from hyperthermophile, Pyrococcus furiosus," *J Mol Biol*, vol. 284, no. 1, pp. 101-24, 1998.

19. Copik, A.J., B.P. Nocek, S.I. Swierczek, S. Ruebush, S.B. Jang, L. Meng, M. D'Souza V, J.W. Peters, B. Bennett and R.C. Holz, "EPR and X-ray crystallographic characterization of the product-bound form of the MnII-loaded methionyl aminopeptidase from Pyrococcus furiosus," *Biochemistry*, vol. 44, no. 1, pp. 121-9, 2005.

20. Watanabe, S., R. Matsumi, T. Arai, H. Atomi, T. Imanaka and K. Miki, "Crystal structures of [NiFe] hydrogenase maturation proteins HypC, HypD, and HypE: insights into cyanation reaction by thiol redox signaling," *Mol Cell*, vol. 27, no. 1, pp. 29-40, 2007.

21. Story, S.V., A.M. Grunden and M.W. Adams, "Characterization of an aminoacylase from the hyperthermophilic archaeon Pyrococcus furiosus," *J Bacteriol*, vol. 183, no. 14, pp. 4259-68, 2001.

22. Wang, X., H.S. Lee, F.J. Sugar, F.E. Jenney, Jr., M.W. Adams and J.H. Prestegard, "PF0610, a novel winged helix-turn-helix variant possessing a rubredoxin-like Zn ribbon motif from the hyperthermophilic archaeon, Pyrococcus furiosus," *Biochemistry*, vol. 46, no. 3, pp. 752-61, 2007.

23. Tatur, J., P.L. Hagedoorn, M.L. Overeijnder and W.R. Hagen, "A highly thermostable ferritin from the hyperthermophilic archaeal anaerobe Pyrococcus furiosus," *Extremophiles*, vol. 10, no. 2, pp. 139-48, 2006.

24. Bryant, F.O. and M.W. Adams, "Characterization of hydrogenase from the hyperthermophilic archaebacterium, Pyrococcus furiosus," *J Biol Chem*, vol. 264, no. 9, pp. 5070-9, 1989.

25. Blamey, J.M. and M.W. Adams, "Purification and characterization of pyruvate ferredoxin oxidoreductase from the hyperthermophilic archaeon *Pyrococcus furiosus*," *Biochimica et Biophysica Acta*, vol. 1161, no. 1, pp. 19-27, 1993.

26. Heider, J., X. Mai and M.W. Adams, "Characterization of 2-ketoisovalerate ferredoxin oxidoreductase, a new and reversible coenzyme A dependent enzyme involved in peptide fermentation by hyperthermophilic archaea," *Journal of Bacteriology*, vol. 178, no. 3, pp. 780-787, 1996.

27. Machielsen, R. and J. van der Oost, "Production and characterization of a thermostable L-threonine dehydrogenase from the hyperthermophilic archaeon Pyrococcus furiosus," *FEBS J*, vol. 273, no. 12, pp. 2722-9, 2006.

28. Hopfner, K.P., L. Craig, G. Moncalian, R.A. Zinkel, T. Usui, B.A. Owen, A. Karcher, B. Henderson, J.L. Bodmer, C.T. McMurray, J.P. Carney, J.H. Petrini and J.A. Tainer, "The Rad50 zinc-hook is a structure joining Mre11 complexes in DNA recombination and repair," *Nature*, vol. 418, no. 6897, pp. 562-6, 2002.

29. Ramsay, B., B. Wiedenheft, M. Allen, G.H. Gauss, C.M. Lawrence, M. Young and T. Douglas, "Dps-like protein from the hyperthermophilic archaeon Pyrococcus furiosus," *J Inorg Biochem*, vol. 100, no. 5-6, pp. 1061-8, 2006.

30. Roy, R., S. Mukund, G.J. Schut, D.M. Dunn, R. Weiss and M.W. Adams, "Purification and molecular characterization of the tungsten-containing formaldehyde ferredoxin oxidoreductase from the hyperthermophilic archaeon Pyrococcus furiosus: the third of a putative five-member tungstoenzyme family," *J Bacteriol*, vol. 181, no. 4, pp. 1171-80, 1999.

31. Jenney, F.E., Jr., M.F. Verhagen, X. Cui and M.W. Adams, "Anaerobic microbes: oxygen detoxification without superoxide dismutase," *Science*, vol. 286, no. 5438, pp. 306-9, 1999.

32. Blake, P.R., J.B. Park, F.O. Bryant, S. Aono, J.K. Magnuson, E. Eccleston, J.B. Howard, M.F. Summers and M.W. Adams, "Determinants of protein hyperthermostability: purification and amino acid sequence of rubredoxin from the hyperthermophilic archaebacterium Pyrococcus furiosus and secondary structure of the zinc adduct by NMR," *Biochemistry*, vol. 30, no. 45, pp. 10885-95, 1991.

33. Weinberg, M.V., F.E. Jenney, Jr., X. Cui and M.W. Adams, "Rubrerythrin from the hyperthermophilic archaeon Pyrococcus furiosus is a rubredoxin-dependent, iron-containing peroxidase," *J Bacteriol*, vol. 186, no. 23, pp. 7888-95, 2004.

34. Ma, K. and M.W. Adams, "Sulfide dehydrogenase from the hyperthermophilic archaeon Pyrococcus furiosus: a new multifunctional enzyme involved in the reduction of elemental sulfur," *J Bacteriol*, vol. 176, no. 21, pp. 6509-17, 1994.

35. Ma, K., R. Weiss and M.W. Adams, "Characterization of hydrogenase II from the hyperthermophilic archaeon Pyrococcus furiosus and assessment of its role in sulfur reduction," *J Bacteriol*, vol. 182, no. 7, pp. 1864-71, 2000.

36. Ghosh, M., A.M. Grunden, D.M. Dunn, R. Weiss and M.W. Adams, "Characterization of native and recombinant forms of an unusual cobalt-dependent proline dipeptidase (prolidase) from the hyperthermophilic archaeon Pyrococcus furiosus," *J Bacteriol*, vol. 180, no. 18, pp. 4781-9, 1998.

37. Sapra, R., M.F. Verhagen and M.W. Adams, "Purification and characterization of a membrane-bound hydrogenase from the hyperthermophilic archaeon Pyrococcus furiosus," *J Bacteriol*, vol. 182, no. 12, pp. 3423-8, 2000.

38. Bevers, L.E., E. Bol, P.L. Hagedoorn and W.R. Hagen, "WOR5, a novel tungsten-containing aldehyde oxidoreductase from Pyrococcus furiosus with a broad substrate Specificity," *J Bacteriol*, vol. 187, no. 20, pp. 7056-61, 2005.

39. Story, S.V., C. Shah, F.E. Jenney, Jr. and M.W. Adams, "Characterization of a novel zinc-containing, lysine-specific aminopeptidase from the hyperthermophilic archaeon Pyrococcus furiosus," *J Bacteriol*, vol. 187, no. 6, pp. 2077-83, 2005.

40. Aono, S., F.O. Bryant and M.W. Adams, "A novel and remarkably thermostable ferredoxin from the hyperthermophilic archaebacterium Pyrococcus furiosus," *J Bacteriol*, vol. 171, no. 6, pp. 3433-9, 1989.

41. Hagen, W.R., P.J. Silva, M.A. Amorim, P.L. Hagedoorn, H. Wassink, H. Haaker and F.T. Robb, "Novel structure and redox chemistry of the prosthetic groups of the iron-sulfur flavoprotein sulfide dehydrogenase from Pyrococcus furiosus; evidence for a [2Fe-2S] cluster with Asp(Cys)3 ligands," *J Biol Inorg Chem*, vol. 5, no. 4, pp. 527-34, 2000.

42. Proudfoot, M., S.A. Sanders, A. Singer, R. Zhang, G. Brown, A. Binkowski, L. Xu, J.A. Lukin, A.G. Murzin, A. Joachimiak, C.H. Arrowsmith, A.M. Edwards, A.V. Savchenko and A.F. Yakunin, "Biochemical and structural characterization of a novel family of cystathionine beta-synthase domain proteins fused to a Zn ribbon-like domain," *J Mol Biol*, vol. 375, no. 1, pp. 301-15, 2008.

43. Roy, R. and M.W. Adams, "Characterization of a fourth tungsten-containing enzyme from the hyperthermophilic archaeon Pyrococcus furiosus," *J Bacteriol*, vol. 184, no. 24, pp. 6952-6, 2002.

44. Sanz-Medel, A., M. Montes-Bayon and M.L.F. Saanchez, "Trace element speciation by ICP-MS in large biomolecules and its potential for proteomics," *Analytical and Bioanalytical Chemistry*, vol. 377, no. 2, pp. 236-247, 2003.

45. Sanz-Medel, A., M. Montes-Bayon, M.D.R.F. de la Campa, J.R. Encinar and J. Bettmer, "Elemental mass spectrometry for quantitative proteomics," *Analytical and Bioanalytical Chemistry*, vol. 390, no. 1, pp. 3-16, 2008.

46. Menon, A.L., F.L. Poole, 2nd, A. Cvetkovic, S.A. Trauger, E. Kalisiak, J.W. Scott, S. Shanmukh, J. Praissman, F.E. Jenney, Jr., W.R. Wikoff, J.V. Apon, G. Siuzdak and M.W. Adams, "Novel multiprotein complexes identified in the hyperthermophilic archaeon Pyrococcus furiosus by non-denaturing fractionation of the native proteome," *Molecular and Cellular Proteomics*, vol. 8, no. 4, pp. 735-751, 2009.

47. Ma, K. and M.W. Adams, "Hydrogenases I and II from Pyrococcus furiosus," *Methods Enzymol*, vol. 331, pp. 208-16, 2001.

48. Schut, G.J., A.L. Menon and M.W. Adams, "2-keto acid oxidoreductases from *Pyrococcus furiosus* and *Thermococcus litoralis*," *Methods in Enzymology*, vol. 331, pp. 144-158, 2001.

49. Mukund, S. and M.W. Adams, "The novel tungsten-iron-sulfur protein of the hyperthermophilic archaebacterium, Pyrococcus furiosus, is an aldehyde ferredoxin oxidoreductase. Evidence for its participation in a unique glycolytic pathway," *J Biol Chem*, vol. 266, no. 22, pp. 14208-16, 1991.

50. Bradford, M.M., "Rapid and sensitive method for quantitation of microgram quantities of protein utilizing principle of protein-dye binding," *Analytical Biochemistry*, vol. 72, no. 1-2, pp. 248-254, 1976.

51. Cvetkovic, A., A.L. Menon, M.P. Thorgersen, J.W. Scott, F.L. Poole, 2nd, F.E. Jenney, Jr., W.A. Lancaster, J.L. Praissman, S. Shanmukh, B.J. Vaccaro, S.A. Trauger, E. Kalisiak, J.V. Apon, G. Siuzdak, S.M. Yannone, J.A. Tainer and M.W. Adams, "Microbial metalloproteomes are largely uncharacterized," *Nature*, vol. 466, no. 7307, pp. 779-782, 2010.

CHAPTER 5

IDENTIFICATION OF A NOVEL NICKEL PROTEIN AND A NOVEL MOLYBDENUM PROTEIN IN THE HYPERTHERMOPHILIC ARCHAEON *PYROCOCCUS FURIOSUS*

Introduction.

A method for identifying novel metalloproteins based on the detection of metals rather than enzyme activity during fraction by column chromatography was recently developed by Cvetkovic et al. [1]. Fractions were also analyzed by peptide mass spectrometry to identify proteins that co-elute with various metals. There are several transition metals that are generally found in metalloproteins. These include V, Mn, Fe, Co, Ni Cu, Zn, Mo and W. The model organism used in the study, *P. furiosus*, has known proteins that associate with a subset of these metals, including Fe, Ni, Co, W, and Zn, but V-, Mn-, and Mo-containing proteins have not been found in *P. furiosus*. However, the method used by Cvetkovic et al. [1] identified several proteins that bind these metals. Interestingly, potential metalloproteins were also identified that bound U and Pb. U and Pb are amongst the most toxic elements to humans and only one naturally-occurring Pb-containing protein, which is involved in detoxification, has been characterized [1]. Purified Pb and U proteins from *P. furiosus* were found to contain their respective metals at substoichiometric levels, and thus these metals are unlikely to be physiologically relevant. Proteins were identified that bound metals at stoichiometric levels, which not only expanded the known number of metalloproteins found in *P. furiosus*, but also the number of metals that are used by proteins in this organism [1].

There were 76 distinct metal-containing peaks that eluted from the chromatography column in the first purification step. These were resolved into 343 peaks when the fractions were purified by a second chromatography step. Nearly half of these (158) could not be identified as known metalloproteins. Those metals that could be assigned to metalloproteins coeluted with proteins that had either been previously characterized as metalloproteins or with proteins that had predicted metal-binding domains. Such a high percentage of proteins that could not be identified as metalloproteins shows that bioinfomatic approaches are insufficient to predict every metalloprotein and that identifying metalloproteins directly from biomass is an effective alternative. Furthermore, identification of additional metalloproteins in this way would be useful in enriching databases with metalloproteins that could consequently help them to become more efficacious in their predictive capabilities.

Herein, we focused efforts on the purification of two metal peaks that could not be assigned to known metalloproteins after the initial fractionation. One peak contained nickel and the other contained molybdenum. Hydrogenases are the only known enzymes in *P. furiosus* that use nickel. They catalyze reactions either involved in the evolution or uptake of hydrogen gas. Therefore, it was interesting that 42 nickel peaks were detected during the fractionation of *P. furiosus* biomass, a number far too great to account for the known nickel enzymes not only in *P. furiosus*, but for all known nickel enzyme characterized from any organism. Even more intriguing is the number of molybdenum peaks identified during fractionation. There were 37 molybdenum peaks identified during the fractionation of *P. furious* biomass. However, a molydoprotein had never been previously characterized or identified in this organism. Thus the identification of both nickel- and molybdenum-containing proteins add new dimensions to understanding of the biochemistry and metabolism of *P. furiosus.*

Materials and methods.

Preparation of anaerobic cell-free extract**.** *P. furiosus* (DSM 3638) was grown under anaerobic, reducing conditions at 90°C in a 600-liter fermenter on maltose and peptides and harvested in late log phase [2]. The procedures for preparing anaerobic cell-free extract and for running the first chromatography step have been described previously [2]. Briefly, 300 g of frozen cells were gently lysed by osmotic shock anaerobically under a continuous flow of argon in 900 ml of 50mM Tris-HCl (pH 8.0), containing 2 mM sodium dithionite as a reductant (DEAE Buffer A) and 0.5 μg/ml DNAse I to reduce viscosity. The cell lysate was centrifuged at 100,000 x g for 2 h at 18 °C, and the clear supernatant, representing the cytoplasmic fraction was immediately loaded onto the first column. This and all subsequent columns were run under anaerobic, reducing conditions where all buffers were degassed and maintained under a positive pressure of argon, and all liquid transfers were made using needles and syringes.

Large scale anaerobic fractionation of cytoplasmic proteins. *First column (c1) fractionation.* All columns were run using an ÄKTA^{TM} basic automated LC system (GE Healthcare). The *P. furiosus* cytoplasmic fraction was diluted 4-fold in Buffer A to reduce the ionic strength and was loaded onto a 10 X 20-cm (1.5-liter) column of DEAE-Sepharose Fast Flow (GE Healthcare) that had been equilibrated previously with the same buffer until the effluent was anaerobic and reducing. Unbound protein was washed off the column with DEAE Buffer A. Bound proteins were eluted using a linear gradient (15 liters) of 0-500 mM NaCl in DEAE Buffer A, and 125-ml anaerobic fractions were collected (120 fractions). Any remaining tightly bound proteins were eluted from the column using 1-liter step gradients of 1, 1.5, and 2 M NaCl respectively, in Buffer A (high salt washes).

Metal analyses. Metal analysis was done by Aleks Cvetkovic and Saratchandra Shanmukh using ICP-MS. A detailed description of the method is can be found elsewhere [3]

Purification of a nickel peak. The nickel profile of the DEAE Sepharose FF column, which is the first purification step, is found in Figure 4.2 of chapter 4. Fractions 28-37 were chosen for further purification. This nickel peak eluted relatively early in the column at an NaCl concentration between 113 mM and 154 mM.

Hydroxyapatite purification of nickel peak. Fractions 28-37 from the previous column were pooled and diluted 20-fold to reduce ionic strength and then loaded onto a 13 X 5-cm (255 mL) column of ceramic hydroxyapatite (BioRad) that had previously been equilibrated with 5mM K-phosphate containing 1mM dithionite (HAP-buffer A) until the column was anaerobic and reducing as determined by the reduction of a solution of methyl viologen. A phosphate gradient from 5 to 500 mM KPO_4 was used in the elution and 50 mL fractions were collected. Fractions that eluted from the hydroxyapatite column between a phosphate concentration of 215 and 235mM were chosen for further purification.

Gel filtration purification of nickel peak. Fractions 31-40 from the previous column were combined and concentrated by 10 kDa MWCO ultrafiltration device (Amicon) to decrease the volume from 200 mL to 5 mL and was loaded onto a 60 X 3.5-cm (577 mL) Superdex 75 (GE Healthcare) gel filtration column that had previously been equilibrated in 50mM Tris HCl, 200mM KCl containing 1mM DT (dithionite) (SX75-buffer A). The fractions (8.5 mL) were collected and the Ni to protein ratios were determined. Protein concentrations were estimated by the method of Bradford [4] using bovine serum albumin as the standard. The SX75 column was calibrated with the standard proteins β-amylase (200 kDa), alcohol dehydrogenase (150

kDa), bovine serum albumin (66 kDa), carbonic andydrase (29 kDa), and cytochrome c (12.5kDa).

SDS-PAGE analysis of SX75 fractions. The fractions were analyzed on a 4-20 % SDS PAGE Criterion® gel (BioRad). The gel was stained by fixing the gel in 25 % (v/v) 2-propanol, 10 % (v/v) acetic acid followed by staining in 0.006 % (w/v) Coomassie Blue R-250, 10 % (v/v) acetic acid. The gel was destained in a mixture of 20 % (v/v)ethanol and 10 %(v/v) acetic acid.

Identification of a novel nickel protein in *P. furiosus*. The procedure for protein identification in solution using high-throughput tandem mass spectrometry (HT-MS/MS) was described previously [2]. The nickel protein was also identified by MALDI of in-gel trypsin digested peptides. SX75 fractions were first separated by SDS-PAGE as described in the previous section. The bands that correspond to fractions which represented the peak with the highest nickel to protein ratio were cut from the gel. The gel bands were processed and digested for 16 h at 37°C according to the manufacturer's protocol. Recombinant porcine trypsin provided with the protocol was used for the in gel digestion (Roche Applied Science). The peptides were purified with C-18 reversed-phase NuTip cartridges according to the manufacturer's instructions (Glygen). The peptides were eluted with 1 μL of a saturated solution of α-cyano-4-hydroxycinnamic acid (Sigma-Aldrich) dissolved in 50 % (v/v) acetonitrile containing 0.1 % (v/v) trifluoroacetic acid (TFA) and spotted onto an MTP 384 Massive MALDI target (Bruker Daltonics) along with 1 μL of ProteoMass Pepide & Protein MALDI-MS Calibration Kit standard (Sigma-Aldrich). The target was analysed using a Bruker Daltonics Autoflex MALDI time-of-flight mass spectrometer in reflectron mode using positive ion detection.

The mass list was generated by the SNAP peak detection algorithm using a signal-to-noise threshold of four following baseline correction of the spectra. Proteins were identified by

searching the mass list against the National Center for Biotechnology Information (NCBI) annotation of the *P. furiosus* genome (NC_003413) using Mascot's Peptide Mass Fingerprint tool (version 2.1, Matrix science). The searches were conducted using a peptide mass tolerance of 1.0, variable modifications of carbamidomethylation (C) and oxidation (M), and a maximum of one missed cleavage. Proteins with a $P< 0.05$ (corresponding to a Mascot protein score greater than 46) were considered significant. Farris Poole, William Lancaster, and Jeremy Praissman carried out metalloprotein bioinformatics analysis. A detailed description of the method can be found elsewhere [3].

Recombinant Protein Expression of PF0086**.** Heterologous expression of PF0086 was carried out in *E. coli* with the addition of either no metal, $NiCl_2$, $CoCl_2$ or $ZnCl_2$ (each 200 μM) to cells growing in NZYCM rich media [3]. The PF0086 open reading frame was amplified from *P. furiosus* genomic DNA using the forward primer 5' GGGAGCTCCATATGACCAGATTGCTATACTATGAAGACGC 3' containing an *NdeI* restriction site and the reverse primer 5' AAGCTCGAGCGGCCGCCTAATCTTCCAGCCATATCTCCAATC 3' containing an *XhoI* restriction site (restriction sites are underlined). The PCR product was digested with the restriction enzymes, Nde1 and Xho1, and inserted into the pET24a(+) vector (Novagen). The sequence of the resulting plasmid, pET24a(+):PF0086, was verified by Sanger sequencing of both strands at the Integrated Biotech Laboratories facility at the University of Georgia. The plasmid was transformed into *E. coli* BL21(DE3) pRIPL, and 1 liter cultures were grown at 37 °C to an OD_{600} of 0.6-0.7. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 0.4 mM and either no metal, $NiCl_2$, $CoCl_2$, or $ZnCl_2$ was added to final a concentration of 200 μM. After a 16 h incubation at 16 °C, cells were harvested by

centrifugation, resuspended in 50 mM Tris, pH 8.0, and lysed with lysozyme. Cell free extracts were prepared by centrifuging the lysed cells at 48,000 x *g* for 20 min. SDS PAGE analysis of the cell free extracts demonstrated the production of a ~25 kDa protein not seen in control cells lacking the recombinant plasmid. To purify PF0086, the cell-extract was heat treated (80 °C for 15 min) and centrifuged at 48,000 x g for 20 minutes. The supernatant was applied to a 5 mL QHP column (GE Healthcare, Piscataway, NJ) equilibrated with 50 mM Tris-HCl, pH 8.0, and the bound proteins were eluted using a linear gradient from 0 to 0.5 M NaCl over 20 column volumes (CVs). The 25 kDa protein eluted from the QHP column at salt concentrations between 0.3 M and 0.4 M NaCl. Those fractions containing the ~25 kDa protein were combined, concentrated, and applied onto a Superdex 75 16/60 column equilibrated with 50 mM Tris-HCl, pH 8.0, containing 200 mM KCl. SDS PAGE analysis was used to identify the fractions containing the ~25 kDa recombinant protein. MALDI-MS analysis confirmed that the major gel band was PF0086, and ICP-MS was used to determine the metal content of the recombinant protein.

Cloning and expression of C717A mutated AlaRS. PF0270 was amplified from a plasmid that had been made during a structural genomics project [5]. The Stratagene QuikChange Lightening Site Directed Mutagenesis kit was used to change cysteine 717 to alanine. The sense primer was 5'-CATTGAAGACTGGGATGTTCAAGCTGCTGGTGGGACTCAC-3' and the antisense primer was 5'-GTGAGTCCCACCAGCAGCTTGAACATCCCAGTCTTCAATG-3'. The underlined bases cause the changes in the amino acid. The sequence of the mutated gene was verified by Sanger sequencing of both strands at the Integrated Biotech Laboratories facility at the University of Georgia. The new plasmid was transformed into *E. coli* BL21(DE3) pRIPL and grown in 1L of 2xYT medium supplemented with kanamycin and chloramphenicol, and

expression was induced when the OD_{600} reached between 0.6 and 0.8 with 0.4mM IPTG. Cells were then grown overnight at 18°C. After harvesting, the cells were lysed and heat treated (15 min. @ 80°C) as a purification step. The SG PCR template caused the new mutated plasmid to have a his tag attached therefore protein was purified over a His column. The heat treated extract was applied to a 5mL Ni-Sepharose 6 Fast flow column (GE Healthcare) column pre-equilibrated with 20mM phosphate, pH7.4, containing 500 mM NaCl and 30mM imidazole (Buffer A). The sample was then eluted from the column using a step gradient. (10 mL step increases of 100mM, 150mM, 200mM, 250mM, 300mM, 400mM, 450mM, and 500 mM imidazole.) The his-tagged protein eluted at 150mM imidazole in a 10 mL volume. C717A AlaRS was then concentrated to 1 mL in an ultrafiltration device having a MWCO of 30kDa. The concentrated C717A AlaRS was loaded onto a 320 mL Superdex 75 that had been equilibrated in buffer (50 mM MOPS, pH 7.0, 200mM KCl, 2 mM DTT, 5 % glycerol).

Purfication of a molybdenum peak. The molybenum profile of the DEAE Sepharose FF column, which is the first purification step, is found in Figure 4.2. Fractions 82-85 were chosen for further purification based on molybdenum and protein content. This molybdenum peak eluted relatively late in the gradient at an NaCl concentration between 325 and 342 mM.

Superdex75 purification of the molybdenum peak. Fractions 82-85 (500 mL) from the previous column were combined and concentrated by ultrafiltration in a device (Amicon) with a MWCO of 10 kDa to decrease the volume to 3 mL. The concentrated pool was loaded onto a (60 X 2.6-cm) 320 mL SX 75 gel filtration column (GE healthcare) that had been equilibrated in buffer A to give the column and anaerobic and reducing environment. Fractions of 5 mL volumes were collected. Fractions eluting between 85 mL and 100 mL were collected and combined. The 320 mL SX75 column was calibrated with the standard proteins β-amylase (200 kDa), alcohol

dehydrogenase (150 kDa), bovine serum album (66 kDa), carbonic andydrase (29 kDa), and cytochrome c (12.5kDa).

Hydroxyapatite purification of the molybdenum peak. Fractions 20-22 from the previous column were pooled and diluted 20-fold to reduce the ionic strength and then loaded onto a (11 X 1.0-cm) 8.6 mL column of ceramic hydroxyapatite (BioRad) that had been previously equilibrated with HAP-buffer A until the column was made anaerobic and reducing. A phosphate gradient from 5 to 500 mM KPO_4 was used in the elution and 2 mL fractions were collected. Fractions that eluted between a phosphate concentration of 147 mM and 178 mM collected for further purification.

Phenyl sepharose purification of the molybdenum peak. Fractions 25-29 (10 mL) from the previous column were pooled and mixed with an equal volume (10 mL) of buffer B [2 M $(NH_4)_2SO_4$, 50 mM Tris-HCl, 1mM DT, pH 7.0]. The mixture was loaded onto a (5 X 0.7-cm) 2 mL HiTrap HP phenyl sepharose column (GE Healthcare) that had been equilibrated in 1M $(NH_4)_2SO_4$ + 50mM Tris-HCl, 1 mM DT, pH 7.0. Proteins were eluted from the column using a gradient from 1M $(NH_4)_2SO_4$ to 0 M $(NH_4)_2SO_4$ and a 10 mL 0 mM $(NH_4)_2SO_4$ wash fraction was collected for further purification.

MonoQ purification of the molybdenum peak. The molybdenum peak eluted in the10 mL low salt wash of the previous column. This fraction was loaded at a 5-fold dilution onto a MonoQ column that had been equilibrated with DEAE-buffer A until the column had an anaerobic and reducing environment. Protein was eluted using an NaCl gradient from 200mM to 400mM over 41 mL. Fractions of 1mL were collected.

Protein identification of purified molybdenum protein**.** MonoQ fractions were run on a 4-20 % Criterion® gel (BioRad) in SDS buffer. The gel was silver stained according to the protocol for

BioRad Silver Stain Plus™. Protein concentrations were determined using the NanoOrange® Protein Quantitation Kit (Invitrogen) and it was determined that 30ng of fraction 36 was loaded into a protein well. Because the protein concentration was too low for the purified protein to be seen by the staining method used to detect the nickel protein and because silver staining was not compatible with MALDI analysis, the fractions were concentrated in a speed vac (Jouan) so that 250 ng of protein could be loaded into a well. The concentrated protein was run on an SDS-PAGE as before except the gel was stained with a higher concentration of Coomassie R-250 (2.5g/L Coomassie Brilliant R-250, 50 % methanol,10 % acetic acid) and destained in 10 % (v/v) acetic acid. MALDI mass spectrometry on the protein extracted from the gel was performed as described above.

Cloning and Expression of PF1972, PF1971, and PF1971-PF1972 dORF**.** The PF1972 open reading frame was amplified from *P. furiosus* genomic DNA using the forward primer 5'-GATAGGTTCCATATGCTTATCGGAGGATGGAAGGC-3' containing an NdeI restriction site and the reverse primer 5'- CCGCTCGAGTTAAGAATACTTCCTGACGATAACTTT-3' containing an XhoI restriction site (restriction sites are underlined). The PCR product was digested with the restriction enzyme, NdeI and XhoI and inserted into the pET21b vector (Novagen). The resulting plasmid was pET21b:PF1972. The PF1971 open reading frame was amplified from *P. furiosus* genomic DNA using the forward primer 5'-GATAGGTTCCATATGGAAAAGGTGGACATCATCCAC-3' containing an NdeI restriction site and the reverse primer 5'-CCGGAGCTCTCAGGCTAAGTAGTGTTTCCTTCTC-3' containing a SacI site (restriction sites are underlined). The PCR product was digested with the restriction enzymes, NdeI and SacI, and inserted into the pET24a(+) vector (Novagen). The sequence of the resulting plasmid was pET24a(+):PF1971. The PF1971/PF1972 double open

reading frame (dORF) was amplified from *P. furiosus* genomic DNA using the forward primer from the PF1971 amplification and the reverse primer for the PF1972 amplification. The PCR product was digested with the enzymes NdeI and SacI and inserted into the pET24a(+) vector resulting in pET24a(+):PF1971/PF1972. The sequences of all resulting plasmids were verified by Sanger sequencing of both strands at the Integrated Biotech Laboratories facility at the University of Georgia.

Overexpression of recombinant enzymes. The dORF plasmid [pET24a(+):PF1971/PF1972] was transformed into *E. coli* BL21(DE3) pRIPL and grown to an A_{600} of 0.6-0.7 in a rich medium containing kanamycin and chloramphenicol. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 0.4mM. After a 16 h incubation at 16°C cells were collected by centrifugation, resuspended in 50mM Tris, 200mM KCl, pH 8.0, and lysed with lysozyme. Cell free extracts were prepared by centrifuging the lysed cells at 48,000*g* for 20 min. SDS-PAGE analysis of the cell-free extracts were compared with extracts from cells that did not contain the plasmid. Heat treatment at 80°C for 15 minutes of the extracts denatured the heat labile *E. coli* proteins. Centrifugation of the heated extracts at 48,000*g* for 20 minutes resulted in extract enriched for the recombinant heat stable *P. furiosus* proteins. pET21b:PF1972 was transformed into *E. coli* BL21(DE3) pRIPL and induced in the same manner as above except ampicillin rather than kanamycin was used as the antibiotic. pET24b:PF1971 and pET24a(+):PF1972 were co-transformed into *E. coli* BL21(DE3) pRIPL and grown in rich media supplemented with ampicillin, kanamycin and chloramphenicol. The culture was induced in the same manner as described above. SDS-PAGE gels (4-20 %) were used to analyze the gel for protein induction. The gel was stained with Coomassie Brilliant Blue R-250 dye.

Results and Discussion.

While 343 metal peaks were detected during the fractionation of *P. furiosus* native biomass, only 7 nickel peaks were detected after the DEAE-FF first purification step, and only one of them could be assigned to hydrogenase, the only known nickel protein in *P. furiosus*. To date, eight nickel dependent enzymes have been described in the scientific literature. These enzymes are urease, Ni-Fe hydrogenase, carbon monoxide dehydrogenase, acetyl-CoA synthase, methyl coenzyme M reductase, glyoxylase, aci-reductone dioxygenase, and methylenediurease [6]. In addition, there are nickel binding proteins that are involved in gene regulation, nickel uptake and efflux, and nickel insertion into enzymes [7]. After further fractionation of the nickel peaks from the first chromatography column, the total number of nickel peaks increased to 42. Of these, 29 could not be assigned to known nickel proteins. Thus, the number of potential nickel proteins detected exceeds the number of nickel proteins known to be present in *P. furiosus*, but also exceeds the number of all known nickel proteins. This observation suggests that there are many unidentified nickel proteins in *P. furiosus*.

The nickel profile of the DEAE fractions is shown in Figure 5.1 along with the profile of the number of unique proteins identified by MS/MS. One major and one minor peak contain subunits of Ni-Fe hydrogenases. However, known or predicted nickel proteins were not detected in the remaining nickel peaks. The most prominent nickel peak was chosen for further purification. It encompassed fractions 31-40 and possessed no detectable hydrogenase subunits by MS/MS analysis. The fractions were pooled and fractionated over a hydroxyapatite column. When this peak was purified using a hydroxyapatite column, it resolved into four nickel peaks designated A, B, C, and D (Figure 5.2). Peak C was chosen for further purification because of its relatively high nickel to protein ratio. Although peak D had a higher ratio it was not chosen

Figure 5.1

Nickel concentration profile and unique peptide profile of *P. furiosus* cell free extract after DEAE fractionation.

ICP-MS and MS/MS data were aligned to indicate the number of protiens that co-elute with nickel peaks during DEAE fractionation. Black arrows indicate nickel peaks in which no known or predicted nickel binding co-elutes. Open arrows indicate nickel peaks that co-elute with a Ni-Fe hydrogenase subunit. The PF numbers designate which Ni-Fe hydrogenase subunits were detected by MS/MS. The black box surrounding the 3rd and 4th arrow from the left indicates the fractions (31-40) that were pooled and fractionated further. PF0894, hydrogenase I alpha subunit; PF1332, hydrogenase II alpha subunit; PF1434 membrane bound hydrogenase alpha subunit.

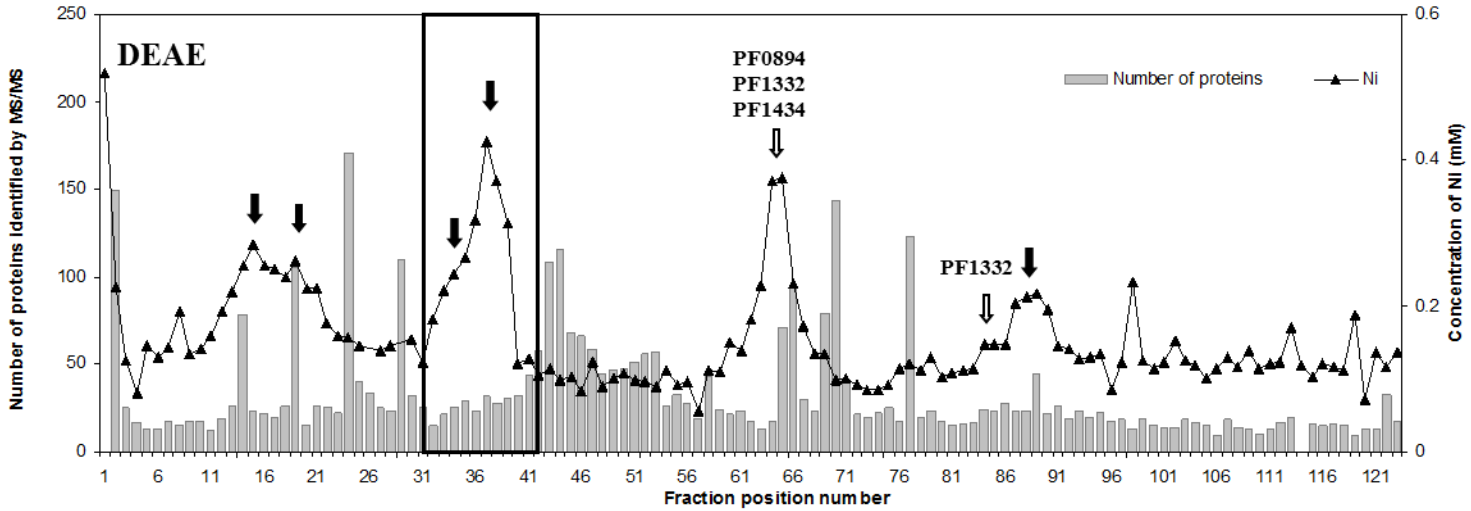
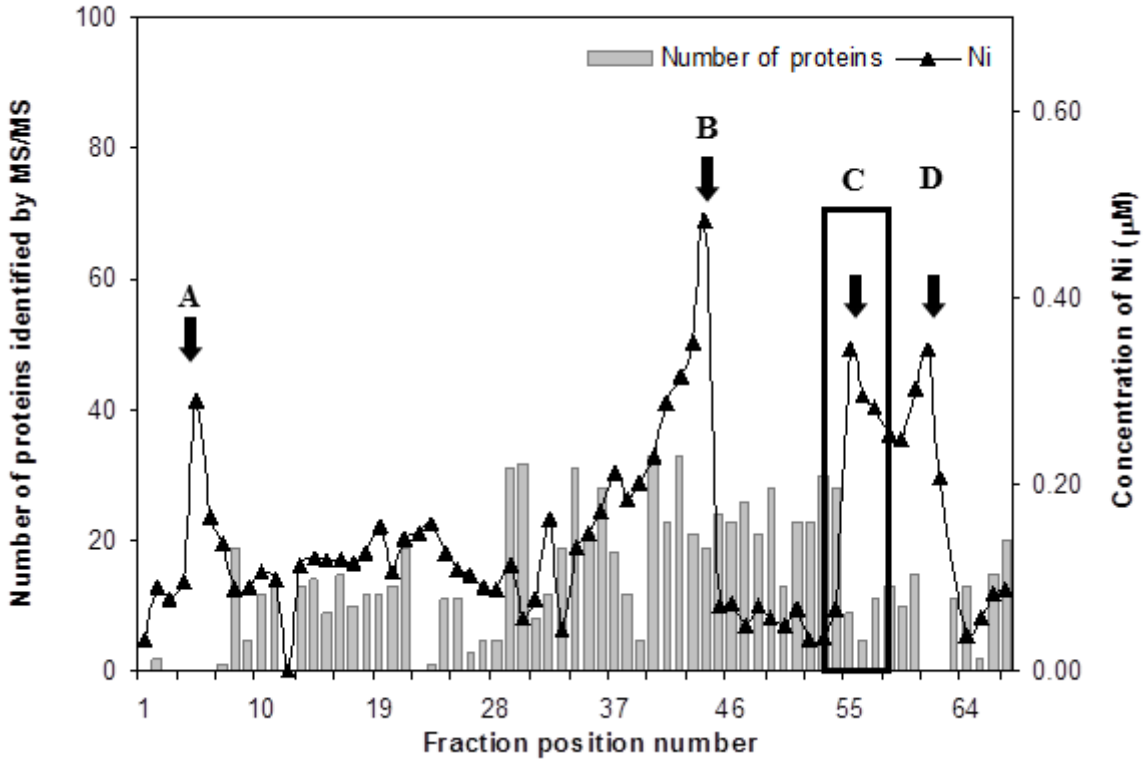

DEAE
PF0894
PF1332
PF1434
PF1332
Number of proteins
Ni
Number of proteins identified by MS/MS
Concentration of Ni (mM)
Fraction position number
0
50
100
150
200
250
0
0.2
0.4
0.6
1
6
11
16
21
26
31
36
41
46
51
56
61
66
71
76
81
86
91
96
101
106
111
116
121

Figure 5.2.

Nickel and peptide elution profile of the DEAE fraction pool after chromatography using a hydroxyapatite column.

ICP-MS and MS/MS data were aligned to indicate the number of proteins that co-elute with nickel peaks during HAP fractionation. The nickel concentration profile show four peaks. The numbers in parentheses indicate the nickel to protein ratio (nmol/mg). Peak D has the highest ratio, the protein concentration was very low (data not shown). Therefore, Peak C was chosen for further purification.


Number of proteins
Ni
A
B
C
D
Number of proteins identified by MS/MS
Concentration of Ni (μM)
Fraction position number
0
20
40
60
80
100
0.00
0.20
0.40
0.60
1
10
19
28
37
46
55
64

because the protein concentrations of its fractions were too low. In addition, peak D included a fraction in which no protein were detected by MS/MS (the reason for this is not known). When peak C was loaded onto a SX75 size exclusion column, it eluted as one major peak (Figure 5.3) that corresponded to a protein with a molecular weight of approximately 25 kDa. In addition to the metal concentration profile obtained with ICP-MS, the protein abundance profile was obtained by plotting the number of unique peptides detected by MS/MS for a particular protein according to the fraction where it is detected. After three chromatography steps, PF0086 was the only protein that had a peptide abundance profile that matched the nickel concentration profile. When the fractions corresponding to the nickel peak were analyzed by SDS-PAGE, they appeared to contain three different protein bands (Figure 5.4). However, after each was extracted from the gel and analysed by MALDI- TOF, it became clear that each band represented the same protein, PF0086 (Figure 5.4). Multiple bands may be the result of degradation or incomplete denaturation. PF0086 is annotated as "alanyl-tRNA synthetase COOH terminus". The homolog from *P. horikoshii* was crystallized (PDB 2E1B) using the recombinant form of the enzyme from *E.coli* and was shown to be monomeric by Fukunga et al. [8]. This agrees with our observation that the predicted size of the *P. furiosus* protein was about 25 kDa, which is about the same size predicted from size exclusion chromatography. The crystal structure showed that the *P. horikoshii* homolog bound one zinc atom. In addition, the primary amino acid sequence of the *P. horikoshii* homolog and the *P. furiosus* proteins indicate that they contain three histidine residues and one cysteine residue that serve as ligands to the zinc atom. Fukunaga et al. analyzed the enzyme activity of the zinc-bound form of the enzyme, so it is not known if the enzyme is active with nickel in the enzyme active site. Furthermore, during purification, the recombinant protein was dialyzed against a buffer containing 0.2mM zinc acetate to ensure that

Figure 5.3.

Purification and analysis of nickel peak C fractions of size exclusion column.

Peak C from the hydroxyapatite column was further fractionated on a Superdex 75 size exclusion column. The nickel concentration profile and MS/MS peptide data were aligned. The peptide profile for PF0086 corresponds to the the nickel concentration profile.

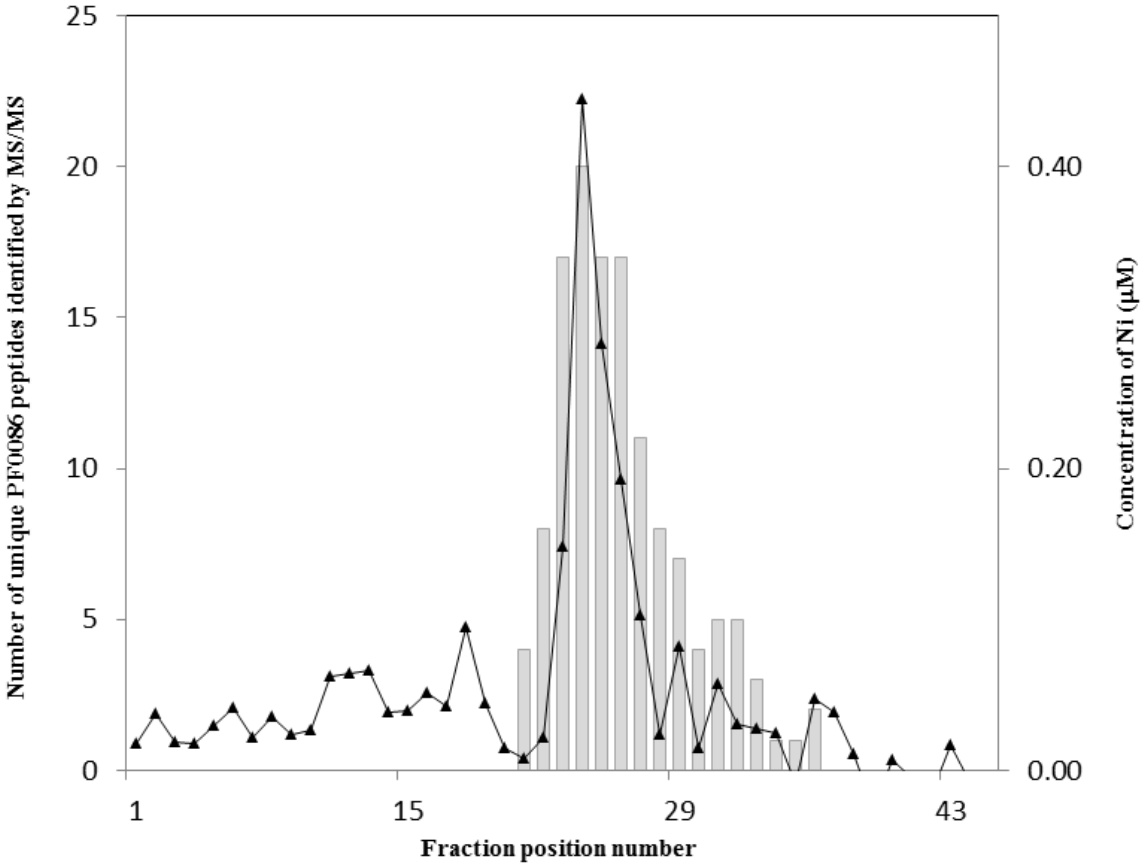
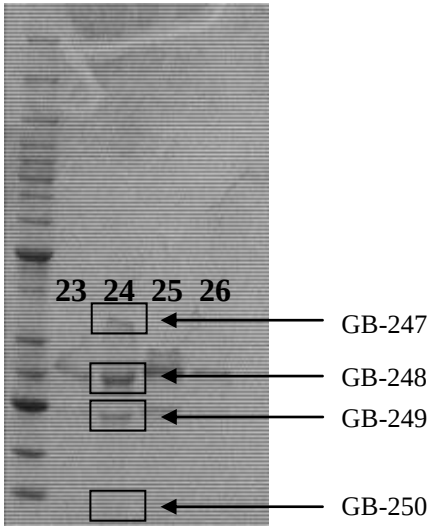

25
20
15
10
5
0
0.40
0.20
0.00
1
15
29
43
Number of unique PF0086 peptides identified by MS/MS
Concentration of Ni (µM)
Fraction position number

Figure 5.4

SDS-gel band MALDI-TOF analysis of the nickel peak fraction.

A. SDS-gel of fractions 23-26. The bands (GB-247 to GB-250) were excised from the gel and analyzed by MALDI-TOF. B. MS/MS data of the SDS gel bands. PF0086 is the only peptide that is common to all bands. This result agrees with the data from the solution MS/MS of the nickel peak fraction.

A



B

GB #	Fr #	PF #	GM Name	MW	PI	% cov	length
GB-247	FR-2950	PF0086	partial alanyl-tRNA synthetase matches COOHterminus	25056	5.7	76.17	214
GB-247	FR-2950	PF0102	conserved hypothetical protein	28106	6.3	43.15	248
GB-247	FR-2950	PF0258	putative endonuclease iv	32411	7.3	14.23	281
GB-247	FR-2950	PF0875	conserved hypothetical protein	33347	9	13.24	287
GB-247	FR-2950	PF0908	conserved hypothetical protein	31947	6.5	12.19	279
GB-247	FR-2950	PF1140	translation initiation factor eIF-2, subunitalpha	31917	7.9	3.64	275
GB-247	FR-2950	PF1589	conserved hypothetical protein	28338	6.9	36.25	251
GB-247	FR-2950	PF1622	n-type ATP pyrophosphatase superfamily	30339	9.3	5.36	261
GB-247	FR-2950	PF1982	conserved hypothetical protein	30612	5.8	24.44	270
GB-248	FR-2950	PF0086	partial alanyl-tRNA synthetase matches COOHterminus	25056	5.7	80.37	214
GB-248	FR-2950	PF1789	hydrolase related to 2-haloalkanoic aciddehalogenase	25237	5.7	7.8	218
GB-248	FR-2950	PF1862	DNA-binding protein	24310	9.1	20.49	205
GB-249	FR-2950	PF0086	partial alanyl-tRNA synthetase matches COOHterminus	25056	5.7	65.42	214
GB-249	FR-2950	PF0284	conserved hypothetical protein	22148	9.1	8.81	193
GB-250	FR-2950	PF0086	partial alanyl-tRNA synthetase matches COOHterminus	25056	5.7	42.06	214
GB-250	FR-2950	PF0440	ribonucleotide reductase	200932	6.1	1.55	1740
GB-250	FR-2950	PF1284	hypothetical protein	6766	9.1	23.73	59
GB-250	FR-2950	PF0222.1	*	7275.6	9.2	15	60

zinc was bound to the enzyme. Hence it is not known what metal the native form of the *P. horikoshii* enzyme contains.

The report of a recombinant PF0086 homolog binding zinc is not consistent with our observation that PF0086, when purified under native conditions from native biomass, is a nickel-containing protein. There are many examples of recombinant proteins that were crystallized with its non-physiological metal [9] and we expect that the PF0086 homolog is one of those cases. To further confirm that PF0086 can bind nickel, the recombinant form of the protein was expressed in *E. coli* in the presence of either 0.2 mM nickel, cobalt, zinc, or no additional metal. The recombinant forms of the protein were purified by heat treatment, anion exchange chromatography and finally size exclusion chromatography. ICP-MS analysis of the purified forms of each recombinant protein preparation showed that PF0086 mostly binds the metal that was added to the growth medium (Figure 5.5). When nickel was added to the medium, PF0086 contained 0.86 ± 0.2 Ni atoms mol^{-1}. It is remarkable that while it appears that *E. coli* inserts whatever metal is available, that *P. furiosus* specifically inserts nickel despite growing in a medium in which zinc is at a ~50- fold higher concentration than nickel. Taken together with the fact that zinc was not detected in the PF0086-containing fractions purified from native biomass, these results are consistent with PF0086 being a nickel-containing protein.

PF0086 is annotated as an alanyl-tRNA synthetase COOH terminus and has four paralogs encoded in the *P. furiosus* genome (Figure 5.6). Each paralog has a homolog encoded in the *P. horikoshii* genome [8]. They are all thought to be involved in editing mischarged tRNAs. PF0086 is a truncated from of alanyl-tRNA synthetase termed Ala X-M where M indicates medium. There are two additional paralogs name AlaX-L (PF0185) and AlaX-S (PF0428), where L stands for large and S for small. The size designation refers to the length of the

Figure 5.5

Metal content of recombinant PF0086 purified from *E. coli* cells grown in rich media or rich media supplemented with either 200 μM Ni, Co, or Zn.

PF0086 was overexpressed in rich media containing either no added metal or added nickel, cobalt, or zinc (each 200 μM). The proteins were purified by heat treatment, anion exchange chromatography and size exclusion chromatography. The fractions from the last purification step were analyzed by ICP-MS. PF0086 bound most to the metal that was added to the media.

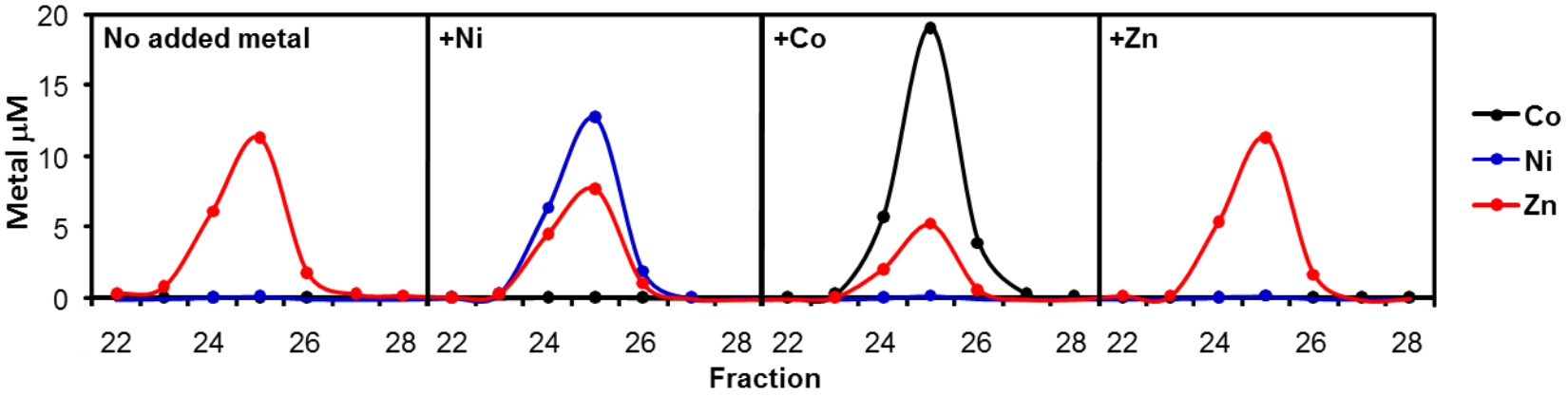
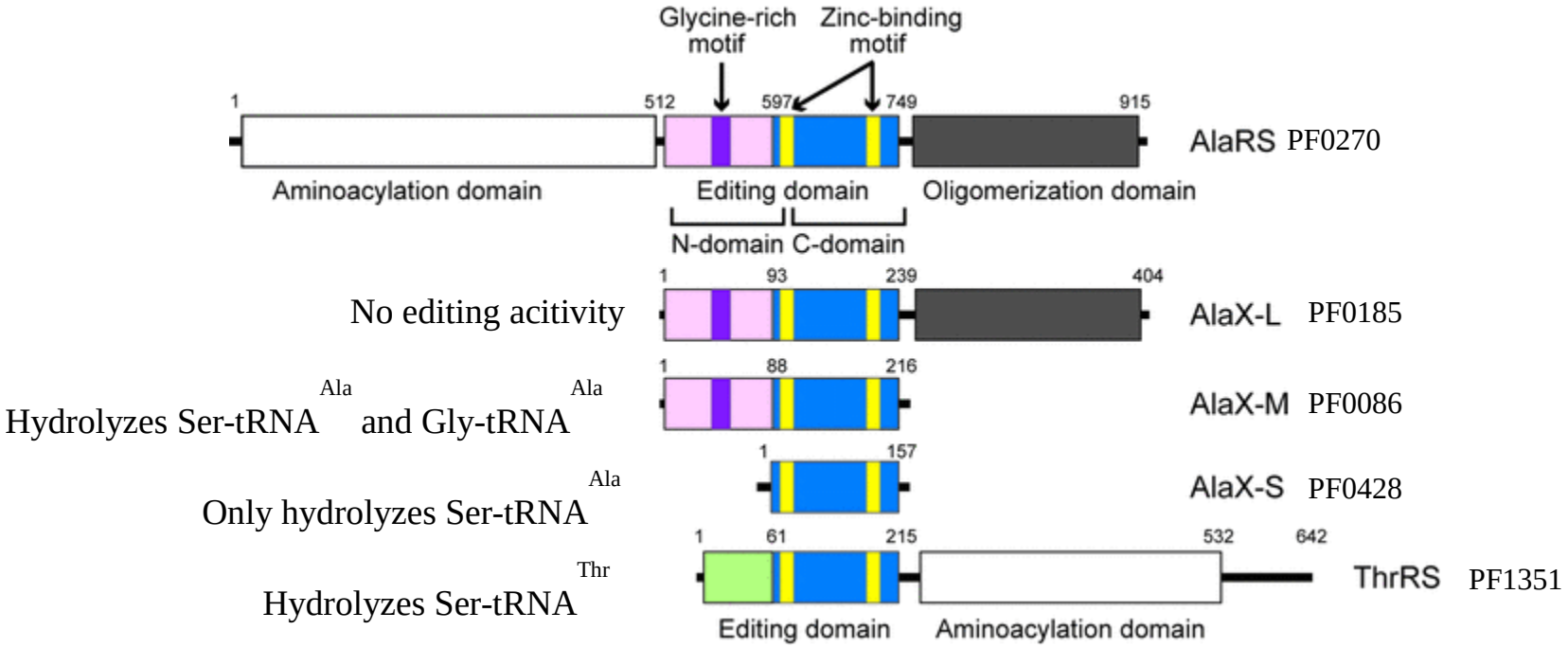

No added metal
+Ni
+Co
+Zn
Metal µM
20
15
10
5
0
22
24
26
28
Fraction
Co
Ni
Zn

Figure 5.6

PF0086 has four paralogs in the *P. furiosus* genome.

PF0086 has four paralogs in the genome and all are predicted to be involved in tRNA editing. They differ in size and function, but all have a conserved metal binding motif that is assumed to bind zinc. Results presented herein suggest that the metal binding domain actually binds nickel. Adapted from [8].


Glycine-rich motif
Zinc-binding motif
1
512
597
749
915
AlaRS PF0270
Aminoacylation domain
Editing domain
Oligomerization domain
N-domain
C-domain
No editing acitivity
1
93
239
404
AlaX-L PF0185
Hydrolyzes Ser-tRNA Ala and Gly-tRNA Ala
1
88
216
AlaX-M PF0086
Only hydrolyzes Ser-tRNA Ala
1
157
AlaX-S PF0428
Hydrolyzes Ser-tRNA Thr
1
61
215
532
642
ThrRS PF1351
Editing domain
Aminoacylation domain

proteins. AlaRS (PF0270) is the full length enzyme and contains domains for charging tRNA with alanine, removing mischarged tRNAs and it also has an oligomerization domain. AlaX- L (PF0185) lacks the aminoacylation and, although it has an editing domain, editing activity has yet to be detected with this enzyme. AlaX-M (PF0086), the enzyme that was purified in this study, lacks both an aminoacylation domain and an oligomerization domain.

AlaX-M (PF0086) can hydrolyze serine and glycine amino acids from mischarged $tRNA^{Ala}$, but cannot add the correct alanine amino acid [8]. AlaX-S (PF0428) is the shortest of the paralogs and does not retain the glycine rich motif although it still has the conserved zinc binding motif. This enzyme can only hydrolyze mischarged serines from $tRNA^{Ala}$. Threonine tRNA synthetase (ThrRS) can remove mischarged serines from $tRNA^{Thr}$ and charge it with the correct amino acid. It also does not have the glycine rich motif but shares the conserved zinc binding domain. The metal binding domain is a common structural component of all of these paralogs and the fractionation data suggests that all of the other paralogs also bind nickel. However, the PF0086 paralogs were not detected (in significant amounts) by MS/MS during the fractionation of *P. furiosus* biomass and so their metal contents remain uncharacterized.

To determine if nickel has any effect on the function of PF0086, it would need to be assayed using a radioactive assay where (^{14}C)-Ser-$tRNA^{Ala}$ or (^{14}C)-Gly-$tRNA^{Ala}$ are the substrates. After the enzyme reaction is spotted on a filter, it is washed with ethanol allowing the free amino acid to wash through the filter, while those which are still bonded to the $tRNA^{Ala}$ are retained on the filter. The amount of radioactivity remaining on the filter indicates the activity level of the enzyme. The substrates are normally made by making a mutated form of AlaRS (C717A AlaRS) which has no editing ability and can only add amino acids. Although we expressed and purified the mutant form of the enzyme unfortunately, our collaborator, Dr. Dieter

Söll (Yale Univ.) was unable to synthesize the substrate after repeated attempts. Hence, even though we had prepared different forms of recombinant PF0086 containing different amounts of nickel, zinc, and/or cobalt, the effect of the metal on activity could not be determined.

The approach used to identify PF1972 as a molybdenum protein was modified from that used previously for PF0086 in that MS/MS was not used during fractionation. Instead the protein was identified solely by MALDI-MS after in-gel protein digestion and extraction of the protein band. In other words, the molybdenum peak was purified on successive chromatography columns until a single protein band was evident on an SDS-gel. The protein was then identified by MS/MS.

PF1972 was one of 18 molybdenum peaks that co-eluted with proteins predicted to bind molybdenum (Figure 5.7). It contained 0.77 ± 0.43 atoms mol^{-1} Mo protein. PF1972 is annotated as an anaerobic ribonucleotide activase while PF1971 is annotated as anaerobic ribonucleotide reductase and is in an operon with PF1972. In both eukaryotes and prokaryotes ribonucleotide reductase synthesize deoxyribonucleotides by the reduction of the 2' hydroxyl group of ribonucleotides. PF1971 is a member of the class of ribonucleotide reductases found in strict and facultative anaerobes which use a catalytically essential glycyl radical in the active site. [10; 11]. PF1972 is the activator responsible for producing the glycyl radical in active site of the reductase.

Anaerobic ribonucleotide reductase activases are "radical-SAM" enzymes and are characterized by a (4Fe-4S) center that is coordinated by three conserved cysteines. However, in hyperthermophiles like *P. furiosus*, there are four conserved cysteine residues in addition to the three found in all anaerobic ribonucleotide reductase activases (Figure 5.8). ICP-MS analysis of the purified *P. furiosus* protein did detect iron in the fraction as well as molybdenum and showed

Figure 5.7

Purification of the novel molybdenum protein PF1972.

The graphs show the elution profile of protein (gray) and molybdenum (blue). The boxed in fractions were pooled and purified through an additional chromatography step. C1, DEAE-Fast Flow anion exchange column; C2, Superdex 75 size exclusion; C3, hydroxyapatite; C4, phenyl sepharose hyrdrophobic interaction; C5, the last step was on a mono Q anion exchange. The purified protein was identified as PF1972 by MALDI-MS. The fractions were analyzed on a denaturing gel and silver stained [3].

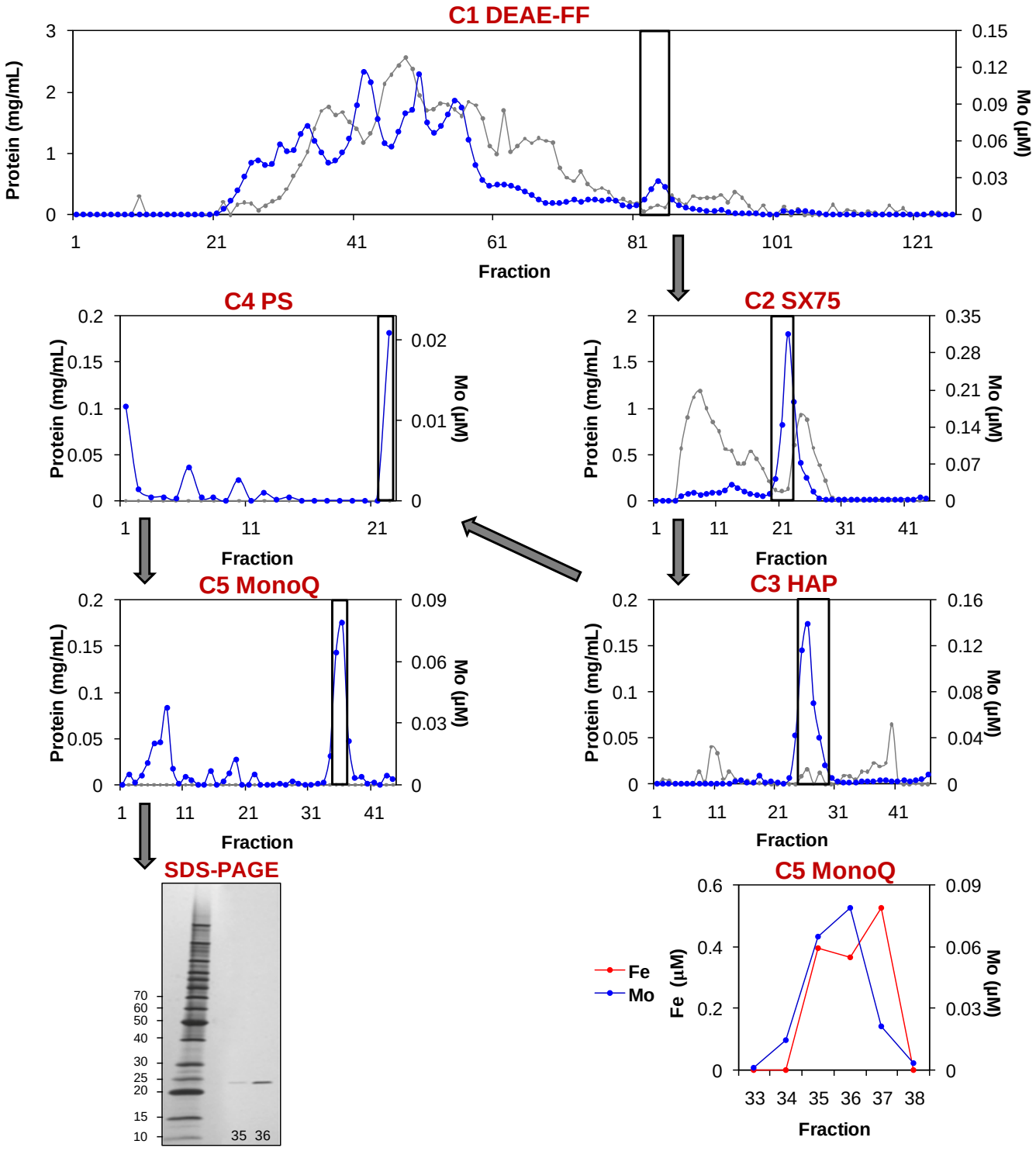
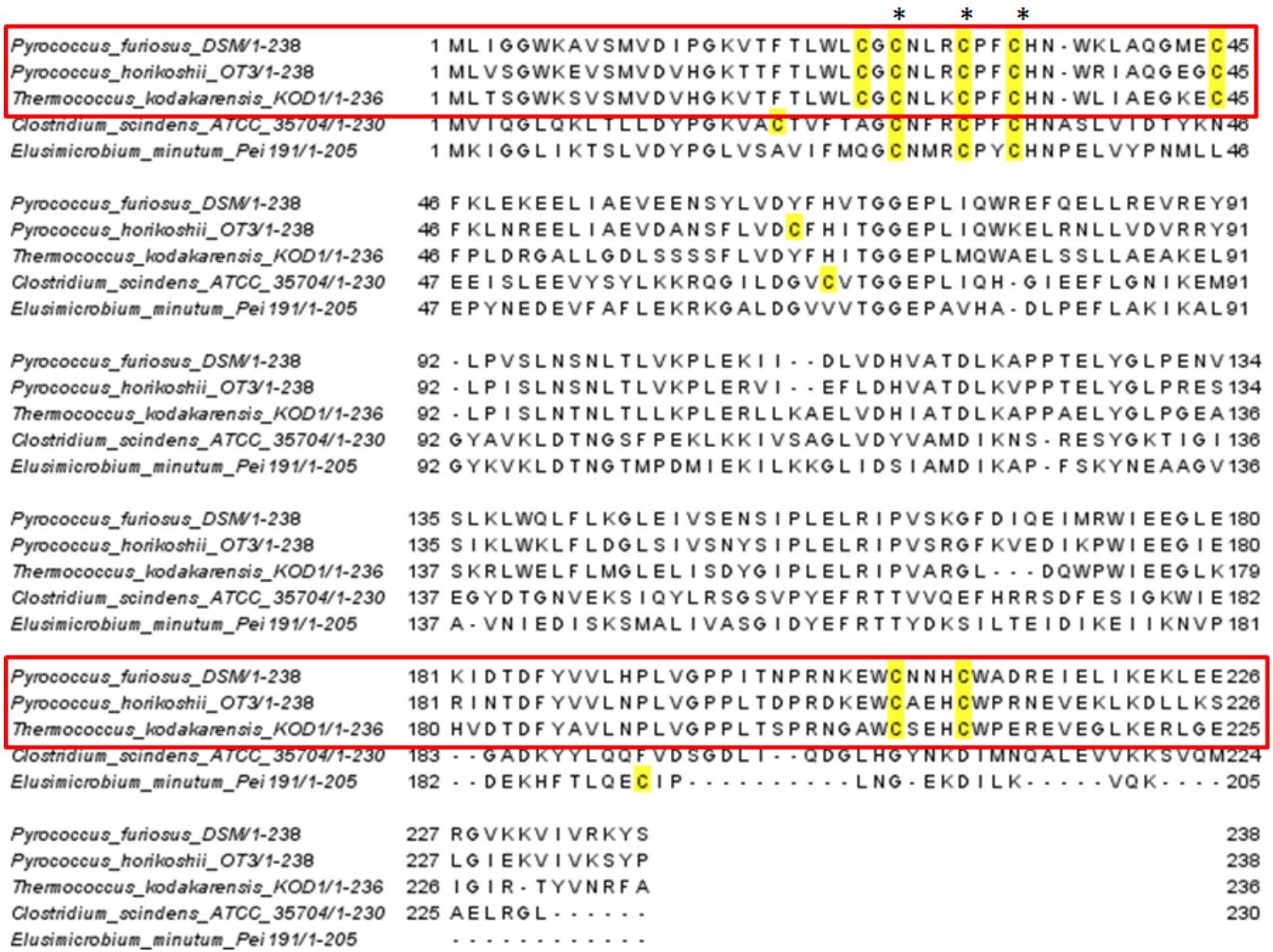

C1 DEAE-FF
Protein (mg/mL)
Mo (μM)
Fraction
C4 PS
C2 SX75
C5 MonoQ
C3 HAP
SDS-PAGE
70
60
50
40
30
25
20
15
10
35 36
C5 MonoQ
Fe
Mo
Fe (μM)

Figure 5.8.

Amino acid alignment of anaerobic ribonucleotide reductase activases.

The alignment shows the three cysteines that are conserved in all RNRs activases. They are indicated by asterisks (*). Four additional cysteines are conserved only in the hyperthermophiles (boxed in red).


Pyrococcus_furiosus_DSM/1-238 1 MLIGGWKAVSMVDIPGKVTFTLWLCGCNLRCPFCHN-WKLAQGMEC 45
Pyrococcus_horikoshii_OT3/1-238 1 MLVSGWKEVSMVDVHGKTTFTLWLCGCNLRCPFCHN-WRIAQGEGC 45
Thermococcus_kodakarensis_KOD1/1-236 1 MLTSGWKSVSMVDVHGKVTFTLWLCGCNLKCPFCHN-WLIAEGKEC 45
Clostridium_scindens_ATCC_35704/1-230 1 MVIQGLQKLTLLDYPGKVACTVFTAGCNFRCPFCHNASLVIDTYKN 46
Elusimicrobium_minutum_Pei191/1-205 1 MKIGGLIKTSLVDYPGLVSAVIFMQGCNMRCPYCHNPELVYPNMLL 46
Pyrococcus_furiosus_DSM/1-238 46 FKLEKEELIAEVEENSYLVDYFHVTGGEPLIQWREFQELLREVREY 91
Pyrococcus_horikoshii_OT3/1-238 46 FKLNREELIAEVDANSFLVDCFHITGGEPLIQWKELRNLLVDVRRY 91
Thermococcus_kodakarensis_KOD1/1-236 46 FPLDRGALLGDLSSSSFLVDYFHITGGEPLMQWAELSSLLAEAKEL 91
Clostridium_scindens_ATCC_35704/1-230 47 EEISLEEVYSYLKKRQGILDGVCVTGGEPLIQH-GIEEFLGNIKEM 91
Elusimicrobium_minutum_Pei191/1-205 47 EPYNEDEVFAFLEKRKGALDGVVVTGGEPAVHA-DLPEFLAKIKAL 91
Pyrococcus_furiosus_DSM/1-238 92 -LPVSLNSNLTLVKPLEKII--DLVDHVATDLKAPPTELYGLPENV 134
Pyrococcus_horikoshii_OT3/1-238 92 -LPISLNSNLTLVKPLERVI--EFLDHVATDLKVPPTELYGLPRES 134
Thermococcus_kodakarensis_KOD1/1-236 92 -LPISLNTNLTLLKPLERLLKAELVDHIATDLKAPPAELYGLPGEA 136
Clostridium_scindens_ATCC_35704/1-230 92 GYAVKLDTNGSFPEKLKKIVSAGLVDYVAMDIKNS-RESYGKTIGI 136
Elusimicrobium_minutum_Pei191/1-205 92 GYKVKLDTNGTMPDMIEKILKKGLIDSIAMDIKAP-FSKYNEAAGV 136
Pyrococcus_furiosus_DSM/1-238 135 SLKLWQLFLKGLEIVSENSIPLELRIPVSKGFDIQEIMRWIEEGLE 180
Pyrococcus_horikoshii_OT3/1-238 135 SIKLWKLFLDGLSIVSNYSIPLELRIPVSRGFKVEDIKPWIEEGIE 180
Thermococcus_kodakarensis_KOD1/1-236 137 SKRLWELFLMGLELISDYGIPLELRIPVARGL---DQWPWIEEGLK 179
Clostridium_scindens_ATCC_35704/1-230 137 EGYDTGNVEKSIQYLRSGSVPYEFRTTVVQEFHRRSDFESIGKWIE 182
Elusimicrobium_minutum_Pei191/1-205 137 A-VNIEDISKSMALIVASGIDYEFRTTYDKSILTEIDIKEIIKNVP 181
Pyrococcus_furiosus_DSM/1-238 181 KIDTDFYVVLHPLVGPPITNPRNKEWCNNHCWADREIELIKEKLEE 226
Pyrococcus_horikoshii_OT3/1-238 181 RINTDFYVVLNPLVGPPLTDPRDKEWCAEHCWPRNEVEKLKDLLKS 226
Thermococcus_kodakarensis_KOD1/1-236 180 HVDTDFYAVLNPLVGPPLTSPRNGAWCSEHCWPEREVEGLKERLGE 225
Clostridium_scindens_ATCC_35704/1-230 183 --GADKYYLQQFVDSGDLI--QDGLHGYNKDIMNQALEVVKKSVQM 224
Elusimicrobium_minutum_Pei191/1-205 182 --DEKHFTLQECIP----------LNG-EKDILK-----VQK---- 205
Pyrococcus_furiosus_DSM/1-238 227 RGVKKVIVRKYS 238
Pyrococcus_horikoshii_OT3/1-238 227 LGIEKVIVKSYP 238
Thermococcus_kodakarensis_KOD1/1-236 226 IGIR-TYVNRFA 236
Clostridium_scindens_ATCC_35704/1-230 225 AELRGL------ 230
Elusimicrobium_minutum_Pei191/1-205 ------------

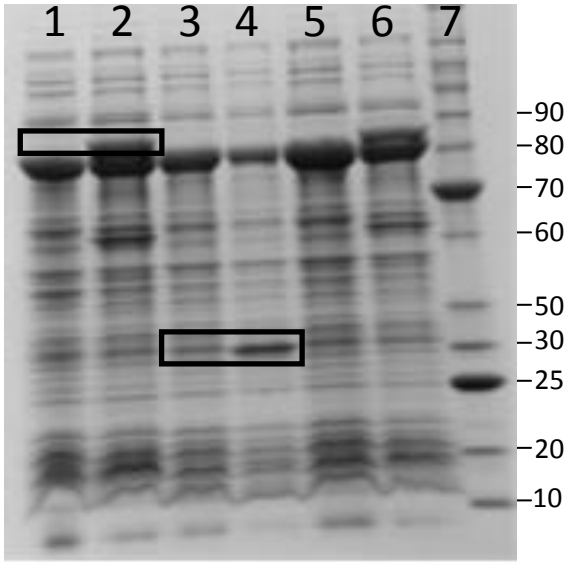
that iron is in a 4 to 1 ratio with molybdenum (Figure 5.7). This ratio would suggest that in addition to the (4Fe-4S) center, one molybdenum atom is bound to the protein. The possibility that the four additional cysteine residues found in the hyperthermophiles coordinate the molybdenum atom is intriguing, since in all other molybdoenzymes, excepting nitrogenase, the metal is coordinated by the sulfur atoms of the organic cofactor molybdopterin (MPT). Furthermore, this would be the first example of a molybdenum enzyme to be found in *P. furiosus*, an archaeon known to depend on tungsten for growth and uses tungsten in the molybdopterin cofactor of several of it enzymes [12].

To confirm the nature of the molybdenum binding, X-ray absorption spectroscopy (XAS) could be used to determine the identity of the molybdenum atom ligands in addition to the geometry of the metal-ligand complex. It would also be interesting to determine the role molybdenum has in the activity of the enzyme. The activity of PF1972 could be measured by the production of dNTP from NTP with a radioactive assay [13]. This would require anaerobic ribonucleotide reductase (PF1971) to be present for activation since PF1971 and PF1972 are predicted to form a complex with $\alpha_2\beta_2$ stoichiometry [14]. However, if AdoMet reductase activity was used to determine activity, the PF1971component would be unnecessary since the methionine produced by the action of the activase alone could be detected after amino acid derivatization and HPLC analysis [11]. Removing the molybdenum or substituting it with another metal may provide some insight into its effect on enzyme activity. However, after five purification steps, the yield of PF1972 protein was only 2.5 μg from 300g of *P. furiosus* cells, an insufficient amount for enzymatic studies. Therefore, genes encoding the activase, along with anaerobic ribonucleotide reductase, were cloned and over-expressed in *E. coli* (Figure 5.9). The PF1971/PF1972 dORF (double open reading frame) was inserted into a pET vector but attempts

Figure 5.9.

Expression of PF1971 and PF1972 in *E. coli* analyzed by SDS-gel electrophoresis.

The subunits of anaerobic ribonucleotide reductase (PF1971 and PF1972) were expressed either separately or together. Shown is the whole cell extracts of the induced cells. The SDS-gel shows extracts of whole *E. coli* cells. Lane 1, PF1971 uninduced; Lane 2, PF1971 induced; Lane 3, PF1972 uninduced; Lane 4, PF1972 induced. Lane 5, PF1971/PF1972 dORF uninduced; Lane 6, PF1971/PF1972 dORF induced; Lanes 7, molecular weight standard. PF1971 has a predicted molecular weight of 70.1 kDa and the predicted molecular weight of PF1972 is 27.6 kDa. Boxes highlight protein bands with sizes corresponding to PF1971 and PF1972.


1
2
3
4
5
6
7
90
80
70
60
50
30
25
20
10

to co-express the proteins from a single plasmid were unsuccessful. PF1971 was expressed, but was insoluble, while PF1972 did not express at all. When co-transformation and co-expression from two separate plasmids was attempted, the same results were obtained (data not shown). PF1971 was also expressed from a plasmid that did not have the PF1972 gene. However, PF1972 was still not overproduced from a plasmid containing only its gene. Although the gel (Figure 5.9) shows a band that appears to be expressed that is roughly the size of PF1972, MALDI of the band indicated that it is an E. *coli* protein and not PF1972 (data not shown).

The results presented herein with PF1972 and PF0086 lend support to the idea that through native purification, information concerning biological molecules can be obtained that could not be discerned from the genome sequence alone. In the case of the PF0086, alanyl-tRNA editing enzyme, the peptide profile of the fractions containing this protein provided the best match to the fractions containing nickel. Experiments with recombinant expression of a *P. horikoshii* homolog suggested that it was a zinc protein, but, the native purification using *P. furiosus* biomass indicated that the enzyme contains nickel. Purification of PF1972, anaerobic ribonucleotide reductase activase, did not include MS/MS analysis during purification. Instead, the protein was purified to homogeneity by following molybdenum and it was identified by MALDI-TOF of peptides extracted from the acrylamide gel. Iron was expected and detected in this protein, however, molybdenum had not been suspected. It was only coincidental that both enzymes purified were already suspected to bind metals and also have known functions. However, metalloproteins were predicted where a metal was not expected and in some cases where a function was not predicted. A study of similar types of proteins provides the greatest potential to expand our knowledge of the roles of metals in biology. It is unfortunate that further analysis of the enzymes purified in this study was not possible, however, the metalloproteomes

of *P. furiosus* and all other organisms provide an almost immeasurable amount of potential targets that can be investigated by similar means.

References

1. Chen, P.R., E.C. Wasinger, J. Zhao, D. van der Lelie, L.X. Chen and C. He, "Spectroscopic insights into lead(II) coordination by the selective lead(II)-binding protein PbrR691," *J Am Chem Soc*, vol. 129, no. 41, pp. 12350-1, 2007.

2. Menon, A.L., F.L. Poole, 2nd, A. Cvetkovic, S.A. Trauger, E. Kalisiak, J.W. Scott, S. Shanmukh, J. Praissman, F.E. Jenney, Jr., W.R. Wikoff, J.V. Apon, G. Siuzdak and M.W. Adams, "Novel multiprotein complexes identified in the hyperthermophilic archaeon Pyrococcus furiosus by non-denaturing fractionation of the native proteome," *Mol Cell Proteomics*, vol. 8, no. 4, pp. 735-751, 2009.

3. Cvetkovic, A., A.L. Menon, M.P. Thorgersen, J.W. Scott, F.L. Poole, 2nd, F.E. Jenney, Jr., W.A. Lancaster, J.L. Praissman, S. Shanmukh, B.J. Vaccaro, S.A. Trauger, E. Kalisiak, J.V. Apon, G. Siuzdak, S.M. Yannone, J.A. Tainer and M.W. Adams, "Microbial metalloproteomes are largely uncharacterized," *Nature*, vol. 466, no. 7307, pp. 779-782, 2010.

4. Bradford, M.M., "Rapid and sensitive method for quantitation of microgram quantities of protein utilizing principle of protein-dye binding," *Anal Biochem*, vol. 72, no. 1-2, pp. 248-254, 1976.

5. Wang, B.C., M.W. Adams, H. Dailey, L. DeLucas, M. Luo, J. Rose, R. Bunzel, T. Dailey, J. Habel, P. Horanyi, F.E. Jenney, Jr., I. Kataeva, H.S. Lee, S. Li, T. Li, D. Lin, Z.J. Liu, C.H. Luan, M. Mayer, L. Nagy, M.G. Newton, J. Ng, F.L. Poole, 2nd, A. Shah, C. Shah, F.J. Sugar and H. Xu, "Protein production and crystallization at SECSG -- an overview," *J Struct Funct Genomics*, vol. 6, no. 2-3, pp. 233-43, 2005.

6. Ragsdale, S.W., "Nickel-based Enzyme Systems," *J Biol Chem*, vol. 284, no. 28, pp. 18571-5, 2009.

7. Mulrooney, S.B. and R.P. Hausinger, "Nickel uptake and utilization by microorganisms," *FEMS Microbiol Rev*, vol. 27, no. 2-3, pp. 239-61, 2003.

8. Fukunaga, R. and S. Yokoyama, "Structure of the AlaX-M trans-editing enzyme from Pyrococcus horikoshii," *Acta Crystallogr D Biol Crystallogr*, vol. 63, no. Pt 3, pp. 390-400, 2007.

9. Eidsness, M.K., S.E. O'Dell, D.M. Kurtz, Jr., R.L. Robson and R.A. Scott, "Expression of a synthetic gene coding for the amino acid sequence of Clostridium pasteurianum rubredoxin," *Protein Eng*, vol. 5, no. 4, pp. 367-71, 1992.

10. Ollagnier, S., E. Mulliez, J. Gaillard, R. Eliasson, M. Fontecave and P. Reichard, "The anaerobic Escherichia coli ribonucleotide reductase. Subunit structure and iron sulfur center," *J Biol Chem*, vol. 271, no. 16, pp. 9410-6, 1996.

11. Mulliez, E., D. Padovani, M. Atta, C. Alcouffe and M. Fontecave, "Activation of class III ribonucleotide reductase by flavodoxin: a protein radical-driven electron transfer to the iron-sulfur center," *Biochemistry*, vol. 40, no. 12, pp. 3730-6, 2001.

12. Johnson, J.L., K.V. Rajagopalan, S. Mukund and M.W. Adams, "Identification of molybdopterin as the organic component of the tungsten cofactor in four enzymes from hyperthermophilic Archaea," *J Biol Chem*, vol. 268, no. 7, pp. 4848-52, 1993.

13. Eliasson, R., E. Pontis, M. Fontecave, C. Gerez, J. Harder, H. Jornvall, M. Krook and P. Reichard, "Characterization of components of the anaerobic ribonucleotide reductase system from Escherichia coli," *J Biol Chem*, vol. 267, no. 35, pp. 25541-7, 1992.

14. Ollagnier, S., C. Meier, E. Mulliez, J. Gaillard, V. Schuenemann, A. Trautwein, T. Mattioli, M. Lutz and M. Fontecave, "Assembly of 2Fe-2S and 4Fe-4S clusters in the anaerobic ribonucleotide reductase from Escherichia coli," *J Am Chem Soc*, vol. 121, no. 27, pp. 6344-6350, 1999.

CHAPTER 6

DISCUSSION AND CONCLUSIONS

One may argue that an underlying premise of this dissertation is that employment of the stereotypical linear scientific method found in high school textbooks would severely limit our progress in science. Many textbooks describe the scientific method as beginning with an observation, followed by a hypothesis or explanation of the observation, experiments to test the hypotheses, experimental results, and finally a conclusion that is drawn from the data collected. Sometimes the data is inconclusive but other times it either supports or contradicts the hypotheses. In the latter case, the hypothesis is rejected and a new hypothesis formulated and tested. Therefore, to perform an experiment without first starting with a seemingly objective observation or with a hypothesis might seem backwards to the novice. However, we were able to make discoveries of new protein-protein and metalloprotein complexes in *P. furiosus* by doing just that.

The main question that was asked when starting this project was, "can we identify novel protein complexes?" In contrast to the textbook version of the scientific method, the study did not begin with an observation per se. The observations were generated *after* experimental analyses of the biomass with MS/MS and ICP-MS. As such, data was collected without being guided by observations or a particular hypothesis in mind. This approach is characteristic of the post-genomics era because now there are many methods that can be used to generate large amounts of data in a short amount of time. It is then the scientists' job to extrapolate workable hypothesis from the data. We used native fractionation and MS/MS analysis of the fractions to

identify 106 potential protein-protein complexes based on co-fractionation of proteins encoded by adjacent genes. The criteria used to identify the complexes left out those complexes that were not encoded by adjacent genes, however the criteria greatly simplified the data so that we could identify those proteins with a high likelihood of forming protein-protein complexes. Of the 106 potential protein complexes that were identified, only 20 were known or expected. Likewise, MS/MS and ICP-MS were used to identify 343 metal peaks resolved during the native fractionation of *P. furiosus* biomass. Of these, 158 co-eluted with proteins that were not predicted to bind a metal. It was from this large data set that the hypotheses that guided the research of this dissertation were generated.

One of the potential 106 novel potential protein-protein complexes was purified through three column chromatography steps as PC-81. The co-eluting proteins of PC-81were encoded by PF1837 and PF1837 ORFs. The PF1837 protein had previously co-purified with PF0532 as ACS2, an $\alpha_2\beta_2$ heterotetrameric enzyme from native biomass. ACS2 and its homolog ACS1 both catalyze terminal steps in the peptide fermentation pathway of *P. furiosus*. ACS1 also is a heterotetramer, but is encoded by PF1540 (α) andPF1787 (β) which are unlinked in the genome. Three additional α subunit homologs are encoded in the genome, but have never been characterized in *P. furiosus*. No additional homologs of the β subunits are encoded in the *P. furiosus* genome. Because the ACS2 β subunit, PF1837, also co-purified with PF1838 as the novel PC-81 complex, we hypothesized that each β subunits was able to form a heterotetrameric enzyme with more than one α subunit. With two β subunits and five α subunits encoded in the genome, we predicted that a total of 10 ACS complexes are used by *P. furiosus* and other members of the Thermococcales which have the same set of genes. The genes of all α and β subunits were cloned and each subunit combination was recombinantly co-expressed in *E. coli*.

Size exclusion chromatography estimated molecular weight of roughly140 kDa for each combination of α and β subunits. This size estimation is in agreement with the $\alpha_2\beta_2$ heterotetrameric stoichiometry found for the native enzymes. Enzyme assays were performed on each recombinant complex to assess their functionality. Acids were used as substrates in the CoA derivative formation direction since acyl-CoAs that account for the degradative pathways of each amino acids were not commercially available. Every combination of subunit formed active enzymes. However, six substrates were not used by any of the recombinant enzymes, which may indicate that those amino acids participate in alternative pathways. Interestingly, acid substrate specificity was dependent only on the α subunit of the complex and was independent of the β subunit. The observation begs the question of the purpose of two β subunit paralogs. Previous work by Mai and Adams [1] using the native enzymes had shown that ACS1 had a greater affinity for guanine nucleotides than adenine nucleotides and ACS2 had a greater affinity for adenine nucleotides over guanine nucleotides, and this hinted at a possible explanation. ACS1 used PF1787 as its β subunit and ACS2 used PF1837 as its β subunit. Bioinformatic analysis of both β subunits designates each as ATP grasp 5 domains which infers that the β subunit contain the ATP binding site. Therefore, we hypothesized that the β subunits confer different nucleotide specificities to the complexes. This hypothesis was tested and supported by our data.

Fukuda et al. [2] described a GTP dependent phosphoenolpyruvate carboxykinase (PEPCK) from *Thermococcus kodakarensis* that was up-regulated when grown in gluconeogenic conditions (pyruvate) and down-regulated when grown in glycolytic conditions (starch). Similarly, the homolog in *P. furiosus* was up-regulated when grown on amino acids and down-regulated when grown on maltose. Therefore, we concluded that when grown under gluconeogenic conditions, the GTP formed by the PF1787 (Gβ) complexes is used by the GTP

dependent PEPCK to drive gluconeogenesis from amino acids. This hypothesis could be tested by varying growth condition of a PF1787 deletion mutant. An observation of good growth under glycolytic conditions, but poor growth under gluconeogenic conditions compared to wild type, would support the hypothesis. Of course, the caveat may be that Gβ ACSs are the only sources of GTP in the cell and other processes such as protein translation may depend on it. However, if there are other sources of GTP that can support GTP dependent processes, a deletion of PF1787 may help to identify what those are.

Alternate β subunits that confer nucleotide specificity to succinyl-CoA synthetase (SCS) have been described in a number of eukaryotes. Similar to the ACSs, a single SCS α subunit can form complexes with two different β subunits which confer specificity to either ATP or GTP. A phylogenetic tree of the β subunits of SCSs found in vertebrates, nematodes, yeast and bacteria suggested that ATP- and GTP- specific β subunits evolved after yeast, but before *C. elegans* [3]. The ACSs in lower life forms appeared to possess only one SCS homolog that was thought to be ATP specific, but were later shown to have broad nucleotide specificity [4].

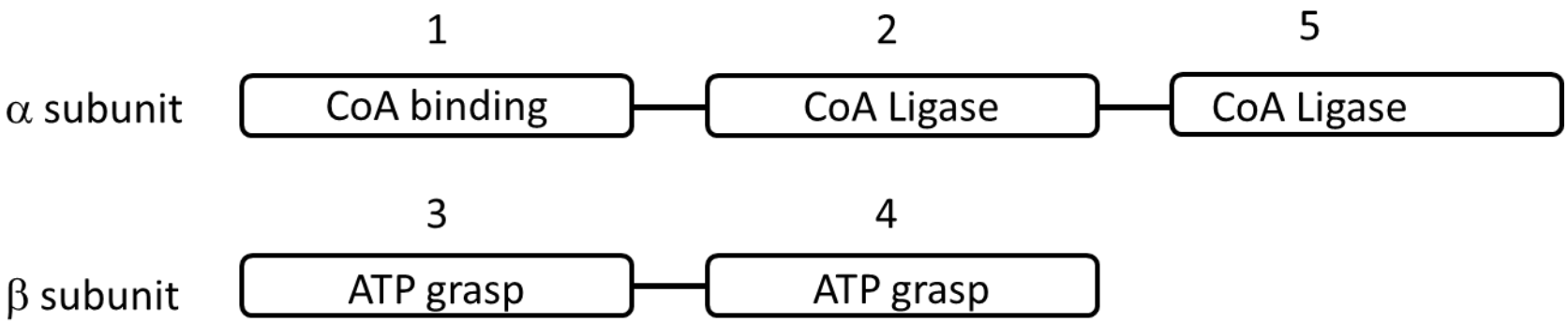
The SCSs and the ACSs of the Thermococcales share the same domains, however, these domains are arranged in different order (Figure 6.1). For example, there is an additional ligase domain in the α subunit and a lack of a ligase domain in the β subunit of *Thermococcus* ACS compared to the SCS. Analysis of the SCS sequences does not reveal any significant matches with *Thermococcus* ACS sequences. Domain rearrangement might be responsible for this result. This may be the reason ACSs from Thermococcales were not included in previous SCS phylogenetic trees [3] In chapter 3, we have included a phylogenetic tree of vertebrate, invertebrate, bacteria, and archaea SCS and ACS β subunits. It is interesting that there appears

Figure 6.1

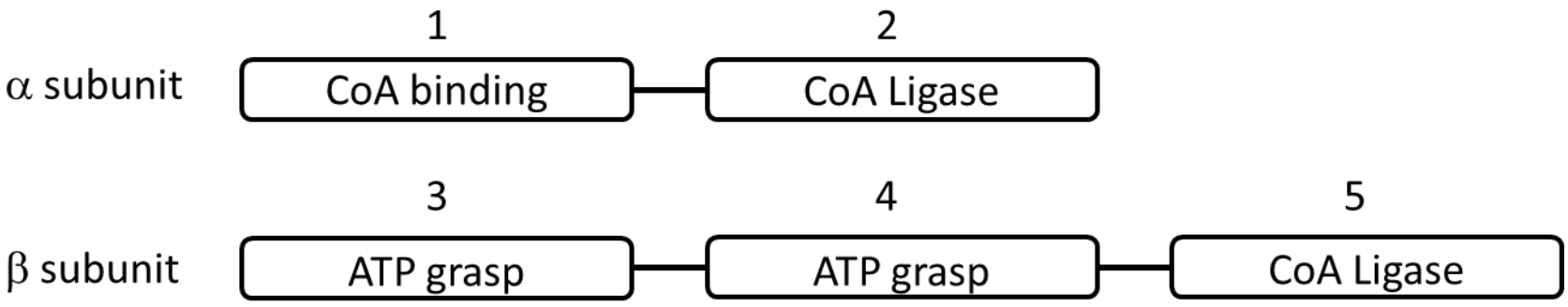
Domain organization of NDP dependent acyl-CoA synthetases (ACSs).

A. Thermococcales ACSs. B. Classical SCS found in bacteria, eukaryotes, and some archaea. Numbers indicate domain order and homologous domains between the ACSs and the classical SCSs. Adapted from [5].

A.


1
2
5
α subunit
CoA binding
CoA Ligase
CoA Ligase
3
4
β subunit
ATP grasp
ATP grasp

B.


1
2
α subunit
CoA binding
CoA Ligase
3
4
5
β subunit
ATP grasp
ATP grasp
CoA Ligase

to be two independent duplications that gave rise to the ATP- and GTP- specific β subunits. This observation suggests that there was selective pressure to compartmentalize ATP and GTP function independently in distantly-related organisms. One possible explanation for this compartmentalization may be due to the presence of an ATP- or GTP- specific enzyme that is dependent on the nucleotide triphosphates (NTP) produced by an NDP-dependent acyl-CoA synthetase. If the sole acyl-CoA synthetase in a cell had broad specificity, there may have been insufficient amounts of a particular NTP produced that was needed by an NTP-dependent enzyme, specific to that particular NTP, to work efficiently.

As a result, a paralog of the β subunit evolved that made more of the NTP dependent enzyme's preferred nucleotide. The acquired specificity of the additional β subunit may have negated the necessity of the original synthetase enzyme to have broad specificity and as a result, it evolved to specifically complement the function of the new ATP- or GTP- specific synthetase. PEPCK is a candidate enzyme that may have contributed to the selective pressure. Since the PEPCK in the Thermococcales, is a GTP dependent enzyme, it would work more efficiently in an environment with an enzyme that specifically makes GTP. It is interesting that in the eukaryote, *Giardia lamblia*, the domains of the homologous enzyme are arranged in the same order as those of the Thermococcales, however, they are fused into one polypeptide. Furthermore, the C-terminal domain of the *G. lamblia* enzyme is homologous to the two β subunit of the Thermococcales proteins. While both *Giardia* and *Thermoccocus* orders use GTP-dependent PEPCKs, only the Thermococcales use the peptide fermentation pathway to generate ATP. Furthermore, the single ACS in *G. lamblia* prefers ATP as a substrate [6]. As such, another method of GTP synthesis occurs in *Giardia* and this method appears to be through the use of GDP as a substrate of PEPCK in an alternative shunt from PEP to pyruvate. In this

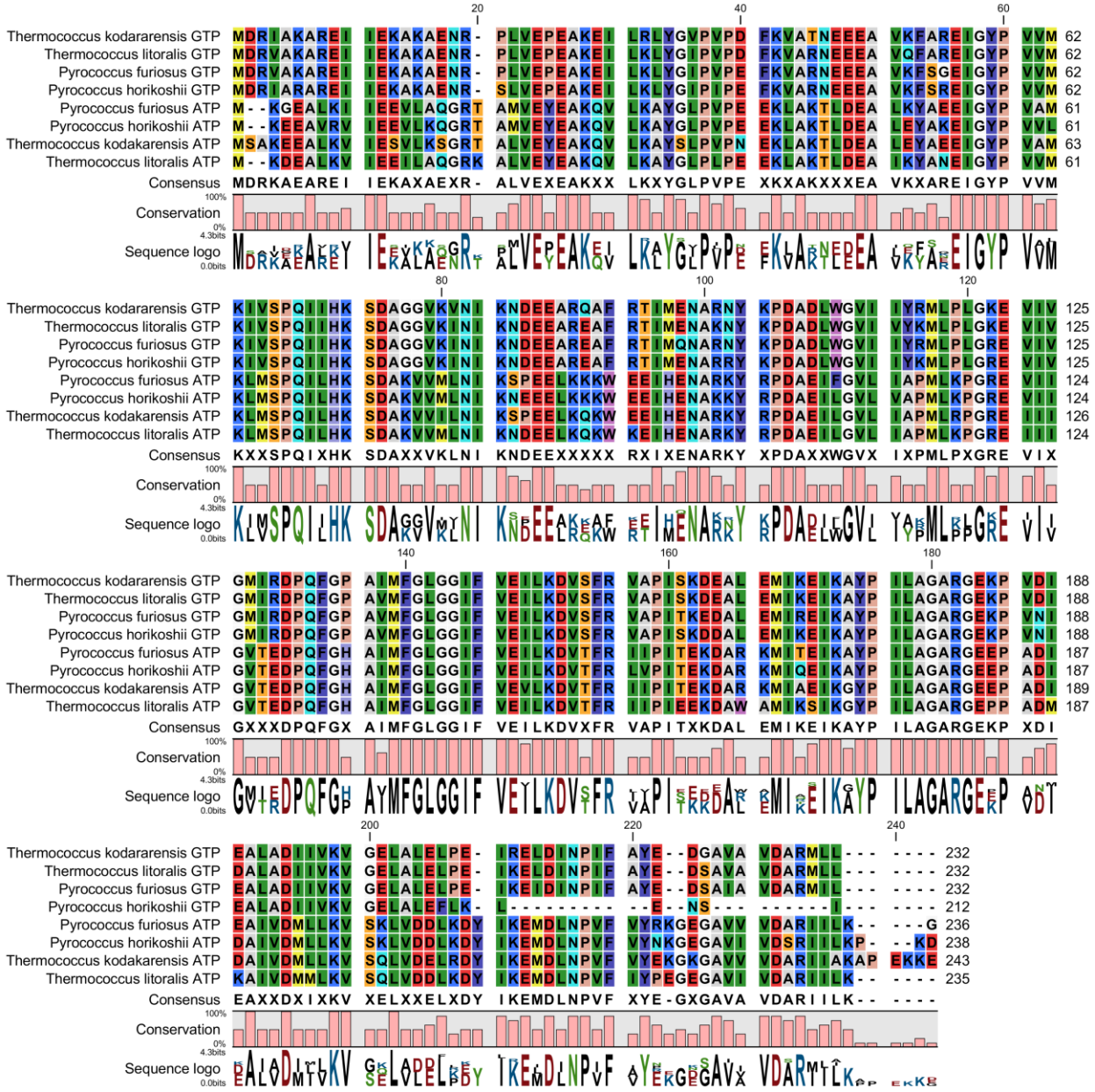
pathway, GTP is produced when PEP is carboxylated by PEPCK to make oxaloacetate which then is reduced to malate. Malate is then decarboxylated to pyruvate [7]**.** *G. lamblia* is an intestinal parasite that uses glucose as its primary carbon source and does not appear to use PEPCK in gluconeogenesis. Thus *Giardia* does not need GTP to drive gluconeogenesis as the Themococcales do, and as such the Thermococcales may have evolved two β subunits to accommodate a need in gluconeogenesis while *Giardia* did not.

The amino acid alignment of the β subunits of some members of the Thermococcales is shown in Figure. 6.2. Residues that are conserved in both the Aβ and Gβ proteins may provide information concerning amino acids that contribute to the overall shape of the subunits, and also their interactions with the α subunits There are also amino acids that are only conserved within the respective Aβ and Gβ subunits. These residues may be involved in determining nucleotide specificity. In particular those residues on the β subunit which are in close spatial proximity to the conserved His^{257} of the α subunit when the enzyme is fully assembled are reasonable candidates. The proposed reaction mechanism of this enzyme shows a phosphorylated conserved $His^{257}\alpha$ which transiently phosphorylates conserved $His^{71}\beta$. $His^{71}\beta$ then phosphorylates ADP to make ATP. A three dimensional model of ACS1 was made by Bräsen et al. [5] using the dimeric crystal structure of the ACS1α homolog in *P. horikoshii* and the crystal structure of the monomeric *P. horikoshii* ACS1β homolog as templates. The model shows that a conserved $Asp^{212}\beta$ is in close proximity to $His^{257}\alpha$ in the fully assembled enzyme, however, this residue is conserved in both Aβ and Gβ thus cannot represent a residue that confers nucleotide specificity. Mutational studied may be necessary to determine which amino acids contribute to nucleotide specificity. It would be interesting if residues which are spatially distal to $His^{257}\alpha$ contribute to nucleotide specificity.

Figure 6.2.

Amino acid alignments Aβ and Gβ subunits of Thermococcales ACSs.

Alignment shows residues that are completely conserved and those that are conserved with in Aβ or Gβ. Those residues that are conserved in all subunits may be important in determining the shape of the protein or may be involved oligomerization while those that are conserved within Aβ or Gβ may determine nucleotide specificity.


Thermococcus kodararensis GTP
Thermococcus litoralis GTP
Pyrococcus furiosus GTP
Pyrococcus horikoshii GTP
Pyrococcus furiosus ATP
Pyrococcus horikoshii ATP
Thermococcus kodakarensis ATP
Thermococcus litoralis ATP
Consensus
Conservation
Sequence logo
100%
0%
4.3bits
0.0bits
20
40
60
80
100
120
140
160
180
200
220
240
62
62
62
62
61
61
63
61
125
125
125
125
124
124
126
124
188
188
188
188
187
187
189
187
232
232
232
212
236
238
243
235

The structural model made by Bräsen included coenzyme A, ADP, and Mg^{2+} as substrates, however it did not include a carboxylic acid [5]. It is interesting that when assaying ACS3 at pH 8.0 the activity towards succinate is only 20 % of that of 3-methylthiopropionate, the substrate representing methionine fermentation (data not shown). However, when the assay was done at a pH of 7.0, SCSs activity toward succinate increased and exceeded its activity with 3-methylthiopropionate. A pH change from 8 to 7 did not significantly alter the activities of any other ACS complex toward any of the other substrates. Succinate was the only dicarboxylic acid substrate used. Although the pKa values of succinate are 4.21 and 5.41, a lowering of the pH may increase the population of the molecule that is protonated at one of the carboxyl groups. Since other substrates have only one negatively charged functional group, the change in pH may cause the succinate to resemble the remainder of the substrates at a pH of 7. This reasoning may also explain why Mai et al. [1] could not detect succinyl CoA synthetase activity when they assayed the fractions at pH8.0. Imanaka et al. [8] were able to detect activity of the native *T. thermokodakarensis* homolog of this enzyme, however they performed their assay at a pH of 6.5 instead of 8. This observation may add to our understanding of substrate binding of the acid and the enzyme reaction mechanism itself.

It is unfortunate that enzyme activity experiments could not be performed with PF0086 and PF1972. We have demonstrated that PF0086 co-purifies with nickel. We also showed that the recombinant enzyme can bind nickel, cobalt or zinc ions when grown in the presence of these metals. Enzyme assays could not be done with PF0086 because the substrate was not available. The mutated full length $tRNA^{Ala}$ synthetase (PF0270 C717A) was made and should have been able to add mischarged serine and glycine amino acid residues while being unable to remove them from $tRNA^{Ala}$. The mutated protein was expressed, purified and sent to our collaborators

who tried using it to synthesize a radioactive substrate. However they were unable to make substrate with this enzyme and attempted to make the substrate with *E. coli* tRNA and a mutated *E coli* $tRNA^{Ala}$ synthetase homolog. Because tRNAs are highly conserved, we thought *E. coli* tRNA could be used as substrate. However, a substrate could not be synthesized with *E. coli* either. It is unclear why this substrate was unable to be generated by our collaborators, who have a great deal of experience with this type of experiment.

Further study of PF1972 was also hindered because it could not be recombinantly expressed in *E. coli*. PF1972 was purified from *P. furiosus* biomass and contained molybdenum and iron. Four conserved cysteine residues exclusively conserved in the hyperthermophilic homolog of PF1972 are the candidate ligands for molybdenum binding. This hypothesis could have been tested if there were sufficient amounts of material to test. Coincidentally, the laboratory of Dr. William Lanzilotta at the University of Georgia was concurrently attempting to express this enzyme and was also unsuccessful in its attempts. Finally, Brian Vaccaro, another graduate student in the Adams' laboratory, attempted to homologously express the protein in *P. furiosus* under the control of a constitutive promoter. Unfortunately his efforts did not meet with success either. It is possible that PF1972 could be over-produced if other promoters are used. For example, our laboratory routinely uses the promoter for the highly-expressed S-layer protein [9]. Therefore, other options are available for the homologous expression of PF1972 in *P. furiosus*. A His-tag was also engineered into PF1972 for homologous expression in *P. furiosus*. It might be worthwhile to see if the absence of this tag enhances expression of the protein. Without the ability to make recombinant forms of the *P. furiosus* enzymes and/or assay their activity the ability of characterizing them is greatly limited. Enzyme activity assays of PF0086 and PF1972 would have been able to demonstrate the effects of various metals on enzyme

activity. Studies have reported that different metals significantly affect the rate of a reaction or even the type of products generated by an enzyme [10; 11]. Such data make it critical to understand which metals bind to enzymes and proteins in their native physiological state.

One may question if the method used in the studies reported in this thesis are the most efficient way to characterize novel protein-protein interactions and the nature of metalloproteins. After all, we were not able to fully characterize PF0086 and PF1972. This might be rephrased as, "is it better to start an experimental project with a hypothesis or with data?" However, this question may be hard to answer since both scenarios have their upsides and downsides. Starting with data may seem less focused and more risky since there is no concrete question being asked. Starting with a question helps one to anticipate future plans and possible caveats. However, because genome-related technology has allowed so much data to be generated, there are more opportunities to discover something novel and interesting. Furthermore, since the data in our studies were gathered from native fractionation, hypotheses can be made that take into account information that is not available from sequence alone.

The use of native biomass fractionation to discover novel metalloproteins and novel protein-protein interactions is not a new concept. However, it is a concept that has increasingly become overshadowed by attempts to use genomics to make predictions concerning the functions of enzymes before they have been expressed and purified. Here we have combined native fractionation with MS/MS and ICP/MS to identify novel protein-protein interactions and novel metalloproteins. It is expected that the information acquired during our experiments can fill some gaps in information about protein-protein complexes and metalloproteins in *P. furiosus*. In many cases this information may be applied to homologous proteins in other organisms. In such

cases, others will be able to make their own hypotheses based on the data gathered from the results of my studies presented herein.

References

1. Mai, X. and M.W. Adams, "Purification and characterization of two reversible and ADP-dependent acetyl coenzyme A synthetases from the hyperthermophilic archaeon *Pyrococcus furiosus*," *J Bacteriol*, vol. 178, no. 20, pp. 5897-5903, 1996.
2. Fukuda, W., T. Fukui, H. Atomi and T. Imanaka, "First characterization of an archaeal GTP-dependent phosphoenolpyruvate carboxykinase from the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1," *J Bacteriol*, vol. 186, no. 14, pp. 4620-4627, 2004.
3. Johnson, J.D., J.G. Mehus, K. Tews, B.I. Milavetz and D.O. Lambeth, "Genetic evidence for the expression of ATP- and GTP-specific succinyl-CoA synthetases in multicellular eucaryotes," *J Biol Chem*, vol. 273, no. 42, pp. 27580-27586, 1998.
4. Weitzman, P.D.J. and H. Jaskowskahodges, "Patterns of Nucleotide Utilization in Bacterial Succinate Thiokinases," *Febs Lett*, vol. 143, no. 2, pp. 237-240, 1982.
5. Bräsen, C., M. Schmidt, J. Grötzinger and P. Schönheit, "Reaction mechanism and structural model of ADP-forming Acetyl-CoA synthetase from the hyperthermophilic archaeon *Pyrococcus furiosus*: evidence for a second active site histidine residue," *J Biol Chem*, vol. 283, no. 22, pp. 15409-15418, 2008.
6. Sànchez, L.B., M.Y. Galperin and M. Müller, "Acetyl-CoA synthetase from the amitochondriate eukaryote *Giardia lamblia* belongs to the newly recognized superfamily of acyl-CoA synthetases (Nucleoside diphosphate-forming)," *J Biol Chem*, vol. 275, no. 8, pp. 5794-5803, 2000.

7. Muller, M., M. Mentel, J.J. van Hellemond, K. Henze, C. Woehle, S.B. Gould, R.Y. Yu, M. van der Giezen, A.G. Tielens and W.F. Martin, "Biochemistry and evolution of anaerobic energy metabolism in eukaryotes," *Microbiol Mol Biol Rev*, vol. 76, no. 2, pp. 444-95, 2012.

8. Shikata, K., T. Fukui, H. Atomi and T. Imanaka, "A novel ADP-forming succinyl-CoA synthetase in *Thermococcus kodakaraensis* structurally related to the archaeal nucleoside diphosphate-forming acetyl-CoA synthetases," *J Biol Chem*, vol. 282, no. 37, pp. 26963-26970, 2007.

9. Hopkins, R.C., J. Sun, F.E. Jenney, Jr., S.K. Chandrayan, P.M. McTernan and M.W. Adams, "Homologous expression of a subcomplex of Pyrococcus furiosus hydrogenase that interacts with pyruvate ferredoxin oxidoreductase," *PLoS One*, vol. 6, no. 10, pp. e26569, 2011.

10. Holmquist, B. and B.L. Vallee, "Metal substitutions and inhibition of thermolysin: spectra of the cobalt enzyme," *J Biol Chem*, vol. 249, no. 14, pp. 4601-7, 1974.

11. Dai, Y., P.C. Wensink and R.H. Abeles, "One protein, two enzymes," *J Biol Chem*, vol. 274, no. 3, pp. 1193-5, 1999.