DEFINING THE GENOME AND METALLOPROTEOME OF MICROORGANISMS

by

WILLIAM ANDREW LANCASTER

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

ABSTRACT

The genome revolution and advent of omic-based approaches have resulted in a dramatic increase in the availability of biological data in just the last few years. The analysis and interpretation of these data can address fundamental questions and reveal unexpected aspects of biology. One such question that is key to the emerging field of synthetic biology is: "What is the minimum complement of genes necessary to sustain life?" Herein it is shown that the answer depends upon the lifestyle of the organism, which in turn directly influences the minimal size of a microbial cell. The 'omics' revolution has also resulted in a dramatic reduction in the cost of DNA sequencing, changing the paradigm from availability of genomes of a few model organisms to a world in which strains, populations and even single cells are completely genotyped. In the current work the sequencing of a recently-isolated genetically tractable strain of *Pyrococcus furiosus* revealed a highly dynamic genome with extensive transposon-mediated rearrangement compared to the model strain. Moreover, disruption of some genes did not result in expected phenotypic changes, indicating greater metabolic redundancy than previously appreciated. It is also shown here that a fundamental aspect of biology that cannot be comprehensively addressed by genome sequence alone is the requirement of specific metal cofactors by proteins for proper structure and function. The coupling of metallomics and

proteomics revealed that the metalloproteomes of microorganisms remain largely uncharacterized. It was shown that *P. furiosus* contains an unexpectedly large number of proteins that bind a metal but were not predicted to be metalloproteins. A bioinformatic analysis of the co-occurrence of metals and proteins during the non-denaturing fractionation of native *P. furiosus* biomass led to the prediction of many potential novel metalloproteins. The study of uptake and utilization of metals by microbes is especially significant in the field of bioremediation. A metallomic and genomic comparison of two model organisms, *Desulfovibrio vulgaris* strain Hildenborough and *Pelosinus fermentans* strain A11 revealed surprising and fundamental differences in metal uptake, providing results that will guide future research into the metal assimilation by these organisms potentially including the characterization of uranium-containing proteins.

INDEX WORDS: *Pyrococcus furiosus*, *Desulfovibrio vulgaris*, *Pelosinus fermentans*, metallomics, proteomics, genome sequencing, minimal cell, bioremediation.

DEFINING THE GENOME AND METALLOPROTEOME OF MICROORGANISMS

by

WILLIAM ANDREW LANCASTER

B.S., Carnegie Mellon University, 2001

B.S., The University of Georgia, 2004

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

2012

DEFINING THE GENOME AND METALLOPROTEOME OF MICROORGANISMS

by

WILLIAM ANDREW LANCASTER

Major Professor:	Michael W. W. Adams
Committee:	Jan Mrázek
	William B. Whitman

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2012

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

The rise of –omics (genomics, proteomics, metallomics, metabolomics, etc.) in recent years has fueled an overwhelming increase in the availability of raw biological data, yet the pace at which such data is being generated by high-throughput methodologies far outstrips the rate at which observations can be appropriately classified, conclusions deduced, and the underlying biological entities characterized. Currently, far fewer resources are required to sequence and assemble the entire genome of a newly discovered microorganism than to functionally or structurally characterize a single novel protein. This will, in all likelihood, soon be true of large multi-billion base pair genomes of multicellular eukaryotes or diverse environmental consortia due to sequencing technology progress driven by efforts to realize affordable personal human genomics (29). The ability to easily sequence entire genomes, not just of newly discovered organisms, but of variant strains with interesting phenotypes or even single cells within a population, will greatly aid the effort to comprehensively characterize the biology of organisms (77). For the foreseeable future, however, a significant proportion of genes in any newly sequenced genome will not have any predicted function (14).

One of the most fundamental questions that remains unanswered is: "What is the minimal set of genes required by a living organism?" This minimal set must include all of the genes necessary for energy utilization, synthesis or acquisition of necessary structural elements, translation of chromosomally encoded information into functional RNA and protein, as well as

replicating this information and producing offspring. The complement of genes present in actual living organisms obviously varies greatly with lifestyle as determined by the environment. Some organisms, typified by parasites, synthesize relatively few components necessary for life from simpler precursors, relying on the host to provide complex molecules such as lipids, amino acids, nucleotides, vitamins and cofactors (5). Free-living organisms may also obtain many complex molecules from the environment, either obligately or only when available thus conserving energy. Others synthesize all or most necessary complex molecules from simpler precursors and are unable to assimilate them even when present in the environment. This strategy is adaptive for oligotrophs inhabiting nutrient poor environments (18). Some streamlined organisms may nevertheless depend on other organisms to create favorable conditions, such as the reduction of hydrogen peroxide (56).

Defining the "minimal genome" is important to understanding basic cellular metabolism by stripping away elements that enhance survival by responding to environmental conditions, evading host defenses and other functions which are not strictly necessary for a model system maintained in a laboratory. Any minimal genome formulated necessarily depends on the desired parameters of the system being considered, and a minimal parasitic genome will be smaller than a minimal photoautotroph. Practical considerations will dictate that any attempts to engineer an organism with a minimal genome will result in an organism that has a reasonable growth rate, is relatively easy to maintain, and is genetically perturbable to probe the interactions among components or to harness the organism to generate useful products in a predictable way minimizing unwanted metabolic byproducts (20). Comparative genomic and proteomic methods have contributed to defining minimal genomes by determining the core proteins and pathways in common among those with a particular lifestyle (25). In Chapter 4 we have analyzed how

minimal genomes vary with biological attributes and also show that while the minimal amount of information required for a replicating organism will vary with its desired characteristics, there are fundamental physical limits to the size of an organism encoding a given amount of information using DNA.

As noted above, a major impact of the omics revolution and the availability of inexpensive DNA sequencing is that the genomes of new strains of model organisms can be rapidly sequenced to provide completely new insights into those strains that would not be apparent in any other way. As part of this thesis I determined and analyzed the genome sequence of a new microbial strain. This work is not presented in the form of a chapter in this thesis (due to joint primary authorship) but the resulting publication is presented in Appendix A (9). This involved a new strain of the hyperthermophilic archaeon, *Pyrococcus furiosus.* Before describing the general background for strain sequencing and the rationale for sequencing this particular strain, the properties of this model organism and its relatives will be described.

PYROCOCCUS FURIOSUS

Pyrococcus furiosus is an anaerobic hyperthermophilic archaeon with an optimal growth temperature of 100ºC. It was first isolated from a shallow marine vent environment near Vulcano, Italy by Karl Stetter in 1986 (23). In addition to its role as a model organism in the study of hyperthermophilc adaptations, industrial applications of its heat-stable enzymes have been explored (41, 42, 89). *P. furiosus* is a heterotroph that utilizes peptides or oligosaccharides as carbon and energy sources. When grown in the presence of sulfur, reductant disposal results in the production of H_2S, while H_2 is produced its absence. The organism's ability to produce H_2 had led to several investigations of its potential in bioenergy production (10, 60, 85).

The *P. furiosus* genome sequence was determined in 2001 and was among the first hyperthermophilic genomes available (74). The genome is typical in size for archaea at 1.9 Mb, with approximately 2100 predicted protein-encoding genes (64). *P. furiosus* is the most well studied model hyperthermophile, and has been extensively analyzed by -omics methods including structural genomics, proteomics, metallomics, transcriptomics and metabolomics. (16, 34, 47, 54, 75, 81, 84, 87, 90).

Complete genomes are available for several closely related species, including *Pyrococcus horikoshii*, *Pyrococcus abyssi*, and *Pyrococcus yayanosii* (13, 38, 40) . A comparison of the three genomes revealed a great degree of genomic plasticity, with extensive chromosomal rearrangements, gene gain and loss and, in the case of *P. furiosus*, extensive transposition and insertion sequence (IS) mediated DNA shuffling (46). *P. horikoshii* and *P. abyssi* are more closely related to each other than either is to *P. furiosus*, and it was proposed that the extensive rearrangements in *P. furiosus* were primarily the result of IS activity (46, 93). *P. furiosus* subspecies *woesei* was independently isolated from the Vulcano environment and initially considered a separate species from *P. furiosus* (92). Although the genome of *P. furiosus woesei* has not been sequenced, the proposal for reclassification was initially made on the basis of high identity among available sequences and extensive colinearity (39). Further analysis by whole genome microarray hybridization revealed a number of genomic segments present in the *P. furiosus* type strain absent in *P. furiosus* woesei (30). Of particular interest was the Mal I operon that encodes an ABC maltose transporter flanked by identical IS, forming a composite transposon which is also found in *Thermococcus litoralis* but not in the other sequenced Pyrococci (30). This discovery provided further evidence of the important role of IS activity in gene exchange and rearrangement in this species.

Until recently, *P. furiosus* suffered from one major limitation as a model organism, namely the lack of system for genetic manipulation. The isolation of a competent strain, *P. furiosus* COM1, has enabled targeted chromosomal gene deletion, heterologous gene expression, native protein tagging and over-expression and transformation with a stable autonomously replicating shuttle vector system (3, 8, 10, 21, 48). A deletion of the *pyrF* (orotidine-5'-phosphate decarboxylase) gene using 5-fluoroorotic acid selection was used in the selection of COM1 resulting in uracil auxotrophy (48). In addition to this selectable marker, strains auxotrophic with respect to tryptophan and agmatine have been generated (3, 22). It has been shown that efficient marker replacement and markerless gene deletion can be achieved in this strain using as few as 40 homologous base pairs on construct flanking regions (22). This ability will greatly aid the manipulation of the *P. furiosus* genome to elucidate the function of individual genes and regulatory networks.

GENOME SEQUENCING OF A NEW STRAIN OF *PYROCOCCUS FURIOSUS*

To better understand the genomic differences between *P. furiosus* COM1 and the reference sequence, second-generation DNA sequencing was employed to obtain a complete genome for this strain. Genomic DNA sequencing was performed using two different technologies and a hybrid assembly approach was used to obtain the complete sequence of COM1. Unpaired 454 sequence was obtained using a quarter plate run on a Roche GS FLX titanium pyrosequencer. In addition, a single lane sequencing was performed on an Illumina GAIIx sequencer with a 2-5 kb mate-pair library. The 454 sequence yielded longer reads (~200bp) than the Illumina (36 bp), while the use of a mate-pair library provided distance constraints on Illumina sequence pairs aiding assembly and scaffolding. Hybrid assembly using

the MIRA assembler and scaffolding using SSPACE resulted in a single chromosomal sequence of the COM1 genome (6, 11). The assembled genome showed a number of rearrangements relative to the published reference sequence (Figure 1.1). IS elements were implicated in the majority of the observed chromosomal rearrangements, consistent with a study of environmental *Pyrococcus* isolates where extensive chromosomal rearrangement and variable distribution of IS elements were observed (88). IS elements, which are present in multiple identical or nearly identical copies, are one type of repetitive DNA that is a serious challenge for second-generation DNA sequence assembly (82). PCR was therefore used to verify the IS mediated rearrangement events indicated by assembly and scaffolding, which revealed multiple possible arrangements of the COM1 genome. Although the original strain from which the NCBI reference sequence was obtained is not available, the strain available in the German collection of microorganisms and cell cultures (DSMZ) was also investigated and revealed rearrangement with respect to the reference sequence. An analysis of the newly sequenced COM1 genome and its metabolic implications is covered in detail in Appendix A (9).

The extensive chromosomal changes and gene disruptions observed in the COM1 genome point to the dramatic evolutionary possibilities of microorganisms that can occur in a short time even in a laboratory environment. The observed changes did not clearly point to a mechanism of the observed competence of this strain, which underscores the current limitations in predicting complex phenotypes from genome sequence alone. The identification of specific genetic differences in this competent strain will facilitate further research using the genetic tools now available to elucidate mechanisms of genetic exchange among organisms in hydrothermal vent environments.

METALLOMICS AND METALLOPROTEOMICS

Another aspect of biology that remains relatively unexplored is the comprehensive characterization of the metallomes of microorganisms (80). The vital biological role of metals has been appreciated for some time, but the emerging field of metallomics is yielding novel insights into the extent of the interaction of organisms with a wide range of metals and metalloids. Natural environments are quite diverse in terms of the metal species present, and their relative abundance over time. This is particularly true of habitats that are geologically very active such as volcanic vents, and those that have been enriched due to human activity. These issues will now be discussed individually in depth, followed by a description of the model microorganisms that are used in the research that is presented in Chapter 3.

Metallomics, the "global analysis of the entirety of metal and metalloid species with a cell or tissue type," is an emerging field which aims to increase our understanding of the diverse role of these critical components of all organisms (80). Metallomics therefore encompasses not only metals that are specifically utilized by cells to perform necessary functions, but also those causing toxic effects through binding to cellular components or catalyzing reactions with toxic byproducts. Cells may have a number of defense mechanisms against such events including active efflux, intracellular sequestration, or chemical modification to less toxic forms. Disruption of metal homeostasis has been implicated either directly or indirectly in a number of human disease including diabetes, cardiovascular disease, cancer, and degenerative neurological disorders such as Parkinson's and Alzheimer's disease (37). In addition, microbial metal metabolism can have profound environmental consequences with implications for human health. These effects can be positive as in the case of the reduction of toxic metals to less soluble forms limiting their migration into water supplies, or the degradation of toxic organic compounds.

Stimulation of these abilities to render environments less toxic is the chief concern of the field of bioremediation. Microbial action can also have negative environmental health effects as in the case of methyl mercury production by microbes by poorly understood mechanisms (28). Methyl mercury is a more potent neurotoxin than other forms that accumulate in aquatic food chains presenting a consumption risk and human cells have no known mechanisms for active excretion (36).

Metal metabolism also plays a key role in microbial action in extreme enviroments, such as those that harbor hyperthermophilic organisms like *P. furiosus* described above. For example, although habitats surrounding hydrothermal vents are often described as extreme, with great variation in temperature, pH, and solutes such as transition metals and sulfide, they may approximate conditions on the early earth under which life arose (53). The availability of soluble and insoluble metals and metal complexes varies greatly in such an environment with local conditions such as pH, temperature, redox potential, sulfide and oxygen concentration. An analysis of the relative mobility of elements in the thermal groundwater of Vulcano Island, where *P. furiosus* was isolated (23), relative to host rock classified Se, Mo, U, W, As, Hg, Ag, Tl, Sb as highly mobile, while Ta, Fe, Co, V, Cr, Pb, Ni, Cd, Mn, Zn and Cu were of intermediate mobility (1). The soluble concentration of these elements vary with the local environment, for example Fe solubility is greatest in acidic anoxic reduced environments, whereas V is more soluble in oxic zones as vanadate oxo-anions, while Cr was stable everywhere (1). While total metal concentrations can be quite high, the formation of metal-sulfide complexes can limit the availability of free metal to organisms in vent environments. It has been demonstrated that *P. furiosus* is able to completely satisfy its Fe requirement by utilizing

dissolved FeS complexes up to at least 3.5 kDa, although the precise mechanism of uptake is not completely understood (15).

The sulfide-mediated limitation in hydrothermal environments of the availability of soluble metal has also been shown to play a protective role for organisms against metal toxicity. A study of three organisms from hydrothermal environments, *Thermococcus fumicolans*, *Pyrococcus* strain GB-D and *Methanocaldococcus jannaschii*, demonstrated that toxic effects of Cu, Co and Zn occurring at concentrations of 1-100 μM were ameliorated by the addition of sulfide to the media (19). The organisms were able to tolerate high total metal concentrations in media when conditions favored dissolved sulfide complexes or solid precipitates, as would be the case in natural environments (19). Hydrothermal vents represent an extreme environment not only in terms of temperature, but also in the concentration of substances toxic to many organisms such as sulfide and metals. By probing adaptations to life at the limits of survivability and particularly their ability to assimilate metals, great insight can potentially be obtained into biological mechanisms across a wide variety of habitats. As described below, part of the research for this thesis involved determining the types of metal assimilated by two hyperthermophilic organisms, *P. furiosus* and *Sulfolobus solfataricus*, which grow in neutrophilic marine vents and in acidophilic freshwater springs, respectively. In addition, a detailed metalloproteomics analysis of *P. furiosus* was carried out (16, 45).

Metalloproteomics focuses on the subset of the metallome that is associated with proteins, which by some estimates includes up to half of the proteome of an organism. Although there are computational tools available to predict whether a protein binds a metal using sequence data alone, they suffer from limitations, and prediction is greatly improved using additional data, such as apo-protein structure which is available only in limited cases (2, 4, 62, 79). Motif based

approaches suffer from high false negative rates since the residues directly involved in metal binding vary from protein to protein (69). In addition, some protein families seem to have strict metal specificity, while other contain members that can bind alternative metals or no metal at all (91).

There are several techniques and technologies that are critical for metalloproteomic studies, which typically involve the separation of metalloproteins and protein complexes followed by identification of the separated components. Separation techniques commonly encountered include 2-D gel electrophoresis, and various types of column chromatography (26). Electrophoretic methods have the advantage of speed and simplicity, but often lack resolution resulting in complex mixtures of proteins complicating identification of specific protein-metal association. Column chromatography has an advantage in resolution because separation can be performed serially, with each step achieving greater separation. Multiple column chemistries can aid in resolving proteins with similar properties, and conditions are readily tailored to preserve interactions by maintaining anaerobic reducing conditions (54). Protein identification in metalloproteomic studies is typically achieved using tandem mass spectrometry, for which there are a great variety of available instrumentation (80). Metal identification is typically achieved using inductively coupled plasma mass spectrometry (ICP-MS) or inductively coupled plasma atomic emission spectrometry (ICP-AES) (51, 80). I was involved in the development of detailed step-by-step protocol for the identification of soluble cytoplasmic metals and metal-associated proteins using column chromatography and ICP-MS. This work is not presented in the form of a chapter in this thesis but the resulting publication is presented in Appendix B (83).

One of the major difficulties encountered in attempts to characterize metalloproteomes on a global basis is the complexity of the samples typically encountered, such as cell-free extracts.

In addition, the abundance of different proteins within a cell varies greatly. Separations achieving low resolution may be sufficient to identify the most abundant metalloproteins, but such an approach may miss proteins that are crucial for cellular function. Part of the research for this thesis involved the development of a completely new experimental approach to identify metalloproteins on a genome-wide basis. This work is not presented in the form of a chapter in this thesis but the resulting publication is presented in Appendix C (16). This approach involved the multistep hierarchical fractionation of native *P. furiosus* biomass with metal (ICP-MS) and protein (MS/MS) data collected for each generated fraction. In addition, a comprehensive characterization and comparison of cytoplasmic metal content for three organisms, *Escherichia coli*, *Sulfolobus solfataricus*, and *P. furiosus* was performed (Figure 1.2). This analysis revealed dramatic differences in the metals assimilated and their relative amounts in the cytoplasm. A peak of metal observed in the elution of a chromatographic separation will have many associated proteins, with the greatest number present at the highest level of the separation hierarchy. In some cases, one or more of the identified proteins will have been previously purified and characterized as binding the observed metal. In most cases, however, no known metalloprotein can explain an observed metal peak. For each such case, it is useful to investigate whether any identified protein potentially binds a metal using comparative techniques. As previously mentioned, the main limitations of such methods using protein sequence alone are the high false negative rate and lack of specificity. The goal of the described procedure, however, was to experimentally find proteins that are not known or predicted to bind metal, so in this case such methods are appropriate. By characterizing metal peaks as unexplained by the presence of a predicted metalloprotein, the number of novel metalloproteins can be estimated (Figure 1.3). The prediction methodology, termed the InterPro-Metal Database (IPM), relies on the use of the

aggregate InterPro (IPR) database and associated software provided by the European Bioinformatics Institute (58). In this approach, predicted protein sequences of the target organism are scanned against motif databases to generate protein to IPR matches. Each IPR entry is annotated with the available functional or structural information for the protein family represented. The IPM procedure classifies each IPR entry for potential metal association by scanning the annotation with regular expressions designed to identify keywords related to metals or metal containing cofactors. The data generated is incorporated into a relational database, which can be associated with the experimental data collected in the metalloproteomic procedure.

The metalloprotemic procedure for the cytoplasmic extract of *P. furiosus* generated metal and protein data for 2,589 chromatographic fractions (45). These data were entered in a relational database which preserved the hierarchical relationships among fractions. A robust data infrastructure enabled querying of the data set across dimensions of identified proteins, identified metals, metal association prediction, and location in the fractionation hierarchy. The processing and analysis of this dataset to predict novel metalloproteins in *P. furiosus* is discussed in detail in Chapter 2 (45). The approach used to characterize cytoplasmic metals and the IPM metalloprotein prediction procedure was extended to two other model organisms, *Desulfovibrio vulgaris* strain Hildenborough, and *Pelosinus fermentans* strain A11. In contrast to the organisms described above, these are mesophilic and inhabit similar environments. They are of great interest as they are models for research in metal metabolism in contaminated environments. The results of their analysis by a metallomics approach are discussed in Chapter 3.

BIOREMEDIATION

While human activity has always resulted in some level of environmental pollution, advances in large scale resource mining, manufacturing and petrochemical processing has resulted in the accumulation of high concentrations of a wide variety of toxins over large areas which present health hazards. These hazards may extend far from the site of contamination through migration of toxins into groundwater and eventually into streams, rivers and the oceans. Large scale production of nuclear weapons has resulted in the acute contamination of many sites around the world, with two of particular interest in the United States, Oak Ridge, Tennessee, and Hanford, Washington. Both sites were established as part of the Manhattan Project. Oak Ridge was the primary site for the separation of fissile ^{235}U (http://www.ornl.gov/info/timeline/40/1940.shtml), while Hanford utilized several reactors for the production of plutonium from 1944 until 1987 (http://www.hanford.gov/page.cfm/HanfordOverview). Both of these sites are subject to remediation under the Comprehensive Environmental Response, Compensation, and Liability Act, commonly known as Superfund (17). Oak Ridge is primarily contaminated with uranium and mercury, while hexavalent chromium is a major contaminant in Hanford soil and groundwater (65, 70). Bioremediation, the use of living organisms to render a contaminated environment less toxic, is an alternative to conventional methods involving removal and processing or sequestering large volumes of material, which is often inefficient, expensive or impractical for highly contaminated sites (78). Several studies have been undertaken to understand microbial community structures, including their change over space and time and the response to nutrient amendment, at the Hanford and Oak Ridge sites (27, 57). A metallomic

comparison of two organisms with great bioremediation potential, *Desulfovibrio vulgaris*, and *Pelosinus fermentans* is presented in chapter 3. A description of these organisms is given below.

DESULFOVIBRIO VULGARIS STRAIN HILDENBOROUGH

Desulfovibrio vulgaris strain Hildenborough (DvH) was first isolated from Wealden clay obtained in Hildenborough, Kent, U.K. in 1946 (66). It is an anaerobic dissimilatory sulfate reducing bacterium (SRB) in the class δ-proteobacteria that oxidizes reduced organic substrates, such as lactate, or molecular hydrogen and reduces sulfate or thiosulfate (7, 32). SRB such as DvH play key roles in the sulfur, carbon and nitrogen cycles and are abundant in anoxic soil, freshwater and marine environments. *Desulfovibrio* species are industrially important because they are responsible for a great deal of corrosion of pipes and equipment encountered in petroleum extraction. The organism is also of great interest in the field of bioremediation, and it is hoped that their ability to reduce toxic metals to less soluble forms may be harnessed to clean contaminated environments (49, 50).

The genome of DvH was sequenced in 2004 (33) and consists of a 3.6 Mb chromosome with 3380 predicted protein coding genes (RefSeq id: NC_002937.3), and an additional 156 coding sequences are present on a 200 Kb megaplasmid (RefSeq id: NC_005863.1) that contains the genes required for nitrogen fixation and the CRISPR system. Both genetic manipulation and transposon-based random mutagenesis systems are available, and various transcriptomic investigations have been undertaken (12, 68). The end products of metabolism are acetate, CO_2, and H_2S (33). DvH can also grow syntrophically without sulfate when co-cultured with a methanogen such as *Methanococcus maripaludis* (86). DvH is known to contain multiple hydrogenases that are present either in the periplasm or bound to the cytoplasmic membrane,

with 90% of hydrogenase activity present in the periplasm (59). DvH hydrogenases include iron-only, nickel-iron (NiFe), and nickel-iron-selenium (NiFeSe) types (33) and DvH was the first obligate anaerobe in which cytochromes were observed (67). A pool of periplasmic cytochromes accepts electrons from the hydrogenases and formate dehydrogenases (63). DvH also contains a menaquinone pool which is reduced by a variety of periplasmic enzymes such as the type II NADH dehydrogenase (43). Electrons stored in these pools in the periplasm are shuttled to the cytoplasm through a variety of membrane bound intermediates for sulfate reduction, which takes place in the cytoplasm (33). Only recently it has been established that the menaquinone pool is directly involved in sulfate reduction via the membrane QmoABC complex (71).

Although DvH is considered a strict anaerobe and has not been shown to use oxygen as a terminal electron acceptor for the purpose of energy conservation, it can enzymatically reduce O_2 via both quinol and cytochrome c oxidases as a defense mechanism to prevent oxidative damage (43). In addition to the expected cytoplasmic enzymes to protect against damage caused by reactive oxygen species such as ruberythrin, rubredoxin-oxygen oxidoreductase and superoxide reductase, there is evidence that DvH up-regulates the production of bacterioferritin to sequester iron in response to oxidative stress to prevent iron-catalyzed Fenton reactions (24).

Metal corrosion is induced by *Desulfovibrio* species both as a result of the production of H_2S, which can be oxidized to H_2SO_4 by sulfide oxidizers after diffusion into oxic zones, and through cathodic depolarization of metal surfaces as part of hydrogen oxidation (61). As with geologically-emitted sulfate in vent systems, endogenously produced sulfide results in the precipitation of a large number of metals in anaerobic soil environments inhabited by DvH. In addition to nonspecific sulfide mediated metal precipitation, DvH has been shown to

enzymatically reduce the toxic metals chromium(VI) and uranium(VI), resulting in less soluble forms. This makes this organism of particular interest in the bioremediation of sites contaminated by mining, industrial and weapons production activities. Although the mechanism of metal reduction is not completely understood, the c3 cytochrome has been shown to function as a Cr(VI) and U(VI) reductase (49, 50). Elucidating the full electron transport chain in DvH both for energy generation and metal reduction remains an area of intense research.

PELOSINUS FERMENTANS

The first representative of the genus *Pelosinus* was cultured from a kaolin sample obtained from the Zhuravlinii Log eluvial deposit near Plast, Russia. *Pelosinus fermentans* (strain R7) was described as a gram negative, iron-reducing, spore-forming heterotroph most closely related to the S*poromusa–Pectinatus–Selenomomas* group in the phylum Firmicutes (76). At the time of its discovery, this group in the family *Veillonellaceae* (formerly *Acidaminococcaceae*) was placed under the class C*lostridia*. Recently, further phylogenetic analysis including additional representatives of this family and other Firmicutes revealed that this group represented a distinct lineage and resulted in the reclassification of *Veillonellaceae* into the order *Selenomonadales* under the newly created class, *Negativicutes (52).* All known members of the *Negativicutes* have Gram-negative cell walls in contrast to other *Firmicutes* which are Gram-positive. The *P. fermentatns* type strain was found to grow on a variety of substrates (lactate, butyrate, pyruvate, malate, succinate, citrate, fumarate, fructose, glucose, mannitol, peptone, Casamino Acids and yeast extract). Shortly after this discovery, a study at Oak Ridge investigating the variation of microbial diversity at various soil depths identified members of the genus *Pelosinus* as a dominant species in lactate-iron enrichment cultures of three of the soil

depth horizons, suggesting a high degree of environmental and metabolic flexibility (31). Another species, *Pelosinus defluvii*, was discovered in groundwater samples from the PetroProcessors of Louisiana, Inc. superfund site in Louisiana, which is contaminated with halogenated solvents (55). This species was also found to reduce Fe when grown on a variety of substrates (casamino acids, fructose, fumarate, glucose, glycerol, pyruvate, and yeast extract), with propionate and acetate as the major products of glucose and glycerol fermentation (55).

Yet another *Pelosinus* strain, UFO1, was isolated from uranium contaminated Oak Ridge sediment and subsequently shown to reduce both Fe and U, and to carry out redox-neutral reactions which result in the formation of U precipitates (72, 73). Reaction of U(VI) with phosphate produces an insoluble mineral precipitate (autunite) and it was demonstrated that UFO1 cells in a phosphate free buffer can mediate this precipitation, presumably using internal phosphate sources. Uranium precipitates were found in association with the plasma membrane, and in stable intracellular granules (72). Uranyl phosphate formation is desirable for bioremediation because the precipitate remains insoluble when the environment is reoxidized, limiting migration into groundwater. A greater understanding of the mechanisms by which *Pelosinus* species carry out these processes is essential for developing strategies for in-situ bioremediation efforts which result in long term sequestration by mineralization of U. One such strategy involved injection of emulsified vegetable oil into a uranium contaminated shallow aquifer at Oak Ridge in which soluble groundwater Uranium concentrations were decreased from initial levels for approximately four months (27). The site was monitored for nine months to track the microbial diversity and geochemical changes. The overall diversity of the community, as determined by 16S rRNA sequencing, decreased, but the overall biomass increased after the

injection. *Pelosinus*-like species became predominant and therefore considered to be "key drivers of subsurface geochemical conditions (27)."

Further support for a major role of *Pelosinus* species in contaminated site came from an experiment carried out to simulate nutrient amendment bioremediation under controlled conditions (57). Groundwater samples from the Hanford 100-H area were used as the innoculum for environmental enrichment cultures. The growths were carried out over 95 days in triplicate glass anaerobic fermentors at 30ºC with constant addition of fresh defined lactate- medium lacking exogenous electron acceptors. Throughout the growth, outflow was analyzed to determine the number and types of organisms present and metabolite concentrations. As with the in situ experiment, overall diversity of the community decreased, and *Pelosinus* species became predominant. Fluorescence-activated cell sorting was used to isolate single cells which were cultured for DNA isolation and 16S rRNA sequencing. A total of 16 isolates were obtained that had greater than 99% 16S rRNA similarity to the *Pelosinus fermentans* R7 type strain. Although all 16 were able to reduce Fe(III), only one isolate, A11, was also able to reduce U(VI), monochromate and dichromate. A draft genome sequence of *Pelosinus fermentans* strain A11 (Pf A11) has been obtained and consists of 5 Mb in 134 contigs with 4758 candidate protein-encoding genes (GenBank accession: AKVM00000000). The prevalence of *Pelosinus* species in highly contaminated sites, and the demonstrated ability to accumulate, mineralize, and reduce metals to less soluble forms make it an attractive target into future research into mechanisms of metal tolerance and detoxification and in the development of practical bioremediation strategies.

RESEARCH OBJECTIVES

The goal of this research is to utilize the advances in genomics, proteomics and metallomics to answer fundamental biological questions. Chapter 4 focuses on the diversity of genome size among currently known organisms with various lifestyles and the physical constraints on cell size imposed by the requirement to encode a given amount of information (44). Chapter 2 focuses on the pursuit and prediction of novel metalloproteins in *P. furiosus* through the analysis of a large data set data generated using a combined proteomic and metallomic investigation of natively fractionated *P. furiosus* biomass (45). The procedure used to generate the data utilized by the predictive analysis, including the computational methods used to predict metalloproteins and thereby identify metal peaks containing potential novel metalloproteins has also been published (16) and this is presented in Appendix C. The results of the sequencing and assembly of the complete genome of a naturally competent strain of *P. furiosus* were also published (9) and are presented in Appendix A. The genetic tools, such as protein tagging and native over-expression (3, 10, 35), which have become available since the discovery of this competent strain, will greatly aid efforts to characterize novel metalloproteins in this species and the complete genome sequence is essential to implementing effective strategies for genetic engineering. Finally, the metallomic analysis of two model organisms, *Desulfovibrio vulgaris* strain Hildenborough and *Pelosinus fermentans* strain A11, which inhabit similar environments with potential utility in bioremediation, is presented in Chapter 3. The analysis revealed fundamental differences in metal uptake, in particular the assimilation of uranium by *P. fermentans* strain A11 that will improve our understanding of microbial interactions with toxic heavy metals in contaminated environments. A detailed protocol of the

metal-based methods used in the analysis of these two model organisms was published (83) and this is given in Appendix B.

REFERENCES

1. **Aiuppa, A., G. Dongarra, G. Capasso, and P. Allard.** 2000. Trace elements in the thermal groundwaters of Vulcano Island (Sicily). Journal of Volcanology and Geothermal Research **98:**189-207.

2. **Babor, M., S. Gerzon, B. Raveh, V. Sobolev, and M. Edelman.** 2008. Prediction of transition metal-binding sites from apo protein structures. Proteins **70:**208-17.

3. **Basen, M., J. Sun, and M. W. Adams.** 2012. Engineering a hyperthermophilic archaeon for temperature-dependent product formation. MBio **3:**e00053-12.

4. **Bertini, I., and G. Cavallaro.** 2010. Bioinformatics in bioinorganic chemistry. Metallomics **2:**39-51.

5. **Blanchard, A., and C. Bebear.** 2012. The evolution of *Mycoplasma genitalium*. Ann N Y Acad Sci **1230:**E61-4.

6. **Boetzer, M., C. V. Henkel, H. J. Jansen, D. Butler, and W. Pirovano.** 2010. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics **27:**578-9.

7. **Brandis, A., and R. K. Thauer.** 1981. Growth of *Desulfovibrio* species on Hydrogen and Sulphate as Sole Energy Source. Journal of General Microbiology **126:**249-252.

8. **Bridger, S. L., S. M. Clarkson, K. Stirrett, M. B. DeBarry, G. L. Lipscomb, G. J. Schut, J. Westpheling, R. A. Scott, and M. W. Adams.** 2011. Deletion strains reveal metabolic roles for key elemental sulfur-responsive proteins in *Pyrococcus furiosus*. J Bacteriol **193:**6498-504.

9. **Bridger, S. L., W. A. Lancaster, F. L. Poole, 2nd, G. J. Schut, and M. W. Adams.** 2012. Genome Sequencing of a Genetically-Tractable *Pyrococcus furiosus* Strain Reveals a Highly Dynamic Genome. J Bacteriol.

10. **Chandrayan, S. K., P. M. McTernan, R. C. Hopkins, J. Sun, F. E. Jenney, Jr., and M. W. Adams.** 2012. Engineering Hyperthermophilic Archaeon *Pyrococcus furiosus* to Overproduce Its Cytoplasmic [NiFe]-Hydrogenase. J Biol Chem **287:**3257-64.

11. **Chevreux, B.** 2005. MIRA: An Automated Genome and EST Assembler. German Cancer Research Center Heidelberg.

12. **Chhabra, S. R., G. Butland, D. A. Elias, J. M. Chandonia, O. Y. Fok, T. R. Juba, A. Gorur, S. Allen, C. M. Leung, K. L. Keller, S. Reveco, G. M. Zane, E. Semkiw, R. Prathapam, B. Gold, M. Singer, M. Ouellet, E. D. Szakal, D. Jorgens, M. N. Price, H. E. Witkowska, H. R. Beller, A. P. Arkin, T. C. Hazen, M. D. Biggin, M. Auer, J. D. Wall, and J. D. Keasling.** 2011. Generalized schemes for high-throughput manipulation of the *Desulfovibrio vulgaris* genome. Appl Environ Microbiol **77:**7595-604.

13. **Chinen, A., I. Uchiyama, and I. Kobayashi.** 2000. Comparison between *Pyrococcus horikoshii* and *Pyrococcus abyssi* genome sequences reveals linkage of restriction-modification genes with large genome polymorphisms. Gene **259:**109-21.

14. **Clark, W. T., and P. Radivojac.** 2011. Analysis of protein function and its prediction from amino acid sequence. Proteins **79:**2086-96.

15. **Clarkson, S. M.** 2011. Effects of Sulfide on Iron Metabolism of the Hyperthermophilic Archaeon *Pyrococcus furiosus*. University of Georgia, Athens.

16. **Cvetkovic, A., A. L. Menon, M. P. Thorgersen, J. W. Scott, F. L. Poole, II, 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. W. Adams.** 2010. Microbial metalloproteomes are largely uncharacterized. Nature (London, U. K.) **466:**779-782.

17. **Deland, M. R.** 1981. Regulatory alert: superfund. Environ Sci Technol **15:**255.

18. **Dewall, M. T., and D. W. Cheng.** 2011. The minimal genome: a metabolic and environmental comparison. Brief Funct Genomics **10:**312-5.

19. **Edgcomb, V. P., S. J. Molyneaux, M. A. Saito, K. Lloyd, S. Boer, C. O. Wirsen, M. S. Atkins, and A. Teske.** 2004. Sulfide ameliorates metal toxicity for deep-sea hydrothermal vent archaea. Appl Environ Microbiol **70:**2551-5.

20. **Elowitz, M., and W. A. Lim.** 2010. Build life to understand it. Nature **468:**889-90.

21. **Farkas, J., D. Chung, M. DeBarry, M. W. Adams, and J. Westpheling.** 2011. Defining components of the chromosomal origin of replication of the hyperthermophilic archaeon *Pyrococcus furiosus* needed for construction of a stable replicating shuttle vector. Appl Environ Microbiol **77:**6343-9.

22. **Farkas, J., K. Stirrett, G. L. Lipscomb, W. Nixon, R. A. Scott, M. W. Adams, and J. Westpheling.** 2012. Recombinogenic Properties of *Pyrococcus furiosus* Strain COM1 Enable Rapid Selection of Targeted Mutants. Appl Environ Microbiol **78:**4669-76.

23. **Fiala, G., and K. Stetter.** 1986. *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100°C. Arch Microbiol **145:**56-61.

24. **Figueiredo, M. C., S. A. Lobo, J. N. Carita, L. S. Nobre, and L. M. Saraiva.** 2012. Bacterioferritin protects the anaerobe *Desulfovibrio vulgaris* Hildenborough against oxygen. Anaerobe.

25. **Fisunov, G. Y., D. G. Alexeev, N. A. Bazaleev, V. G. Ladygina, M. A. Galyamina, I. G. Kondratov, N. A. Zhukova, M. V. Serebryakova, I. A. Demina, and V. M. Govorun.** 2011. Core proteome of the minimal cell: comparative proteomics of three mollicute species. PLoS One **6:**e21964.

26. **Garcia, J. S., C. S. Magalhaes, and M. A. Arruda.** 2006. Trends in metal-binding and metalloprotein analysis. Talanta **69:**1-15.

27. **Gihring, T. M., G. Zhang, C. C. Brandt, S. C. Brooks, J. H. Campbell, S. Carroll, C. S. Criddle, S. J. Green, P. Jardine, J. E. Kostka, K. Lowe, T. L. Mehlhorn, W. Overholt, D. B. Watson, Z. Yang, W. M. Wu, and C. W. Schadt.** 2011. A limited microbial consortium is responsible for extended bioreduction of uranium in a contaminated aquifer. Appl Environ Microbiol **77:**5955-65.

28. **Gilmour, C. C., D. A. Elias, A. M. Kucken, S. D. Brown, A. V. Palumbo, C. W. Schadt, and J. D. Wall.** 2011. The Sulfate-reducing bacterium *Desulfovibrio desulfuricans* ND132 as a model for understanding bacterial mercury methylation. Appl Environ Microbiol **77:**3938-51.

29. **Gonzaga-Jauregui, C., J. R. Lupski, and R. A. Gibbs.** 2012. Human genome sequencing in health and disease. Annu Rev Med **63:**35-61.

30. **Hamilton-Brehm, S. D., G. J. Schut, and M. W. Adams.** 2005. Metabolic and evolutionary relationships among *Pyrococcus* Species: genetic exchange within a hydrothermal vent environment. J Bacteriol **187:**7492-9.

31. **Hansel, C. M., S. Fendorf, P. M. Jardine, and C. A. Francis.** 2008. Changes in bacterial and archaeal community structure and functional diversity along a geochemically variable soil profile. Appl Environ Microbiol **74:**1620-33.

32. **Haschke, R. H., and L. L. Campbell.** 1971. Thiosulfate reductase of *Desulfovibrio vulgaris*. J Bacteriol **106:**603-7.

33. **Heidelberg, J. F., R. Seshadri, S. A. Haveman, C. L. Hemme, I. T. Paulsen, J. F. Kolonay, J. A. Eisen, N. Ward, B. Methe, L. M. Brinkac, S. C. Daugherty, R. T. Deboy, R. J. Dodson, A. S. Durkin, R. Madupu, W. C. Nelson, S. A. Sullivan, D. Fouts, D. H. Haft, J. Selengut, J. D. Peterson, T. M. Davidsen, N. Zafar, L. Zhou, D. Radune, G. Dimitrov, M. Hance, K. Tran, H. Khouri, J. Gill, T. R. Utterback, T. V. Feldblyum, J. D. Wall, G. Voordouw, and C. M. Fraser.** 2004. The genome sequence of the anaerobic, sulfate-reducing bacterium *Desulfovibrio vulgaris* Hildenborough. Nat Biotechnol **22:**554-9.

34. **Holden, J. F., F. L. Poole Ii, S. L. Tollaksen, C. S. Giometti, H. Lim, J. R. Yates Iii, and M. W. Adams.** 2001. Identification of membrane proteins in the hyperthermophilic archaeon *Pyrococcus furiosus* using proteomics and prediction programs. Comp Funct Genomics **2:**275-88.

35. **Hopkins, R. C., J. Sun, F. E. Jenney, Jr., S. K. Chandrayan, P. M. McTernan, and M. W. Adams.** 2011. Homologous expression of a subcomplex of *Pyrococcus furiosus* hydrogenase that interacts with pyruvate ferredoxin oxidoreductase. PLoS One **6:**e26569, Oct 24 (epub).

36. **Houston, M. C.** 2011. Role of mercury toxicity in hypertension, cardiovascular disease, and stroke. J Clin Hypertens (Greenwich) **13:**621-7.

37. **Jomova, K., and M. Valko.** 2011. Advances in metal-induced oxidative stress and human disease. Toxicology **283:**65-87.

38. **Jun, X., L. Lupeng, X. Minjuan, P. Oger, W. Fengping, M. Jebbar, and X. Xiang.** 2011. Complete genome sequence of the obligate piezophilic hyperthermophilic archaeon *Pyrococcus yayanosii* CH1. J Bacteriol **193:**4297-8.

39. **Kanoksilapatham, W., J. M. Gonzalez, D. L. Maeder, J. DiRuggiero, and F. T. Robb.** 2004. A proposal to rename the hyperthermophile *Pyrococcus woesei* as *Pyrococcus furiosus* subsp. *woesei*. Archaea **1:**277-83.

40. **Kawarabayasi, Y., M. Sawada, H. Horikawa, Y. Haikawa, Y. Hino, S. Yamamoto, M. Sekine, S. Baba, H. Kosugi, A. Hosoyama, Y. Nagai, M. Sakai, K. Ogura, R. Otsuka, H. Nakazawa, M. Takamiya, Y. Ohfuku, T. Funahashi, T. Tanaka, Y. Kudoh, J. Yamazaki, N. Kushida, A. Oguchi, K. Aoki, and H. Kikuchi.** 1998. Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium, *Pyrococcus horikoshii* OT3. DNA Res **5:**55-76.

41. **Kengen, S. W., E. J. Luesink, A. J. Stams, and A. J. Zehnder.** 1993. Purification and characterization of an extremely thermostable beta-glucosidase from the hyperthermophilic archaeon *Pyrococcus furiosus*. Eur J Biochem **213:**305-12.

42. **Kim, H. W., and K. Ishikawa.** 2010. Complete saccharification of cellulose at high temperature using endocellulase and beta-glucosidase from *Pyrococcus* sp. J Microbiol Biotechnol **20:**889-92.

43. **Lamrabet, O., L. Pieulle, C. Aubert, F. Mouhamar, P. Stocker, A. Dolla, and G. Brasseur.** 2011. Oxygen reduction in the strict anaerobe *Desulfovibrio vulgaris* Hildenborough: characterization of two membrane-bound oxygen reductases. Microbiology **157:**2720-32.

44. **Lancaster, W. A., and M. W. W. Adams.** 2011. The Influence of Environment and Metabolic Capacity on the Size of a Microorganism, p. 93-103. *In* P. L. Luisi and P. Stano (ed.), The Minimal Cell: The Biophysics of Cell Compartment and the Origin of Cell Functionality. Springer Netherlands.

45. **Lancaster, W. A., J. L. Praissman, F. L. Poole, 2nd, A. Cvetkovic, A. L. Menon, J. W. Scott, F. E. Jenney, Jr., M. P. Thorgersen, E. Kalisiak, J. V. Apon, S. A. Trauger, G. Siuzdak, J. A. Tainer, and M. W. Adams.** 2011. A computational framework for proteome-wide pursuit and prediction of metalloproteins using ICP-MS and MS/MS data. BMC Bioinformatics **12:**64.

46. **Lecompte, O., R. Ripp, V. Puzos-Barbe, S. Duprat, R. Heilig, J. Dietrich, J. C. Thierry, and O. Poch.** 2001. Genome evolution at the genus level: comparison of three complete genomes of hyperthermophilic archaea. Genome Res **11:**981-93.

47. **Lee, A. M., J. R. Sevinsky, J. L. Bundy, A. M. Grunden, and J. L. Stephenson, Jr.** 2009. Proteomics of *Pyrococcus furiosus*, a hyperthermophilic archaeon refractory to traditional methods. J Proteome Res **8:**3844-51.

48. **Lipscomb, G. L., K. Stirrett, G. J. Schut, F. Yang, F. E. Jenney, Jr., R. A. Scott, M. W. Adams, and J. Westpheling.** 2011. Natural competence in the hyperthermophilic

archaeon *Pyrococcus furiosus* facilitates genetic manipulation: construction of markerless deletions of genes encoding the two cytoplasmic hydrogenases. Appl Environ Microbiol **77:**2232-8.

49. **Lovley, D. R., and E. J. Phillips.** 1994. Reduction of Chromate by Desulfovibrio vulgaris and Its c(3) Cytochrome. Appl Environ Microbiol **60:**726-8.

50. **Lovley, D. R., P. K. Widman, J. C. Woodward, and E. J. Phillips.** 1993. Reduction of uranium by cytochrome c3 of *Desulfovibrio vulgaris*. Appl Environ Microbiol **59:**3572-6.

51. **Manley, S. A., and J. Gailer.** 2009. Analysis of the plasma metalloproteome by SEC-ICP-AES: bridging proteomics and metabolomics. Expert Rev Proteomics **6:**251-65.

52. **Marchandin, H., C. Teyssier, J. Campos, H. Jean-Pierre, F. Roger, B. Gay, J. P. Carlier, and E. Jumas-Bilak.** 2010. Negativicoccus succinicivorans gen. nov., sp. nov., isolated from human clinical samples, emended description of the family *Veillonellaceae* and description of *Negativicutes* classis nov., *Selenomonadales* ord. nov. and *Acidaminococcaceae* fam. nov. in the bacterial phylum Firmicutes. Int J Syst Evol Microbiol **60:**1271-9.

53. **Martin, W., J. Baross, D. Kelley, and M. J. Russell.** 2008. Hydrothermal vents and the origin of life. Nat Rev Microbiol **6:**805-14.

54. **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.** 2009. Novel multiprotein complexes identified in the hyperthermophilic archaeon *Pyrococcus furiosus* by non-denaturing fractionation of the native proteome. Mol Cell Proteomics **8:**735-51.

55. **Moe, W. M., R. E. Stebbing, J. U. Rao, K. S. Bowman, M. F. Nobre, M. S. da Costa, and F. A. Rainey.** 2011. *Pelosinus defluvii* sp. nov., isolated from chlorinated solvent contaminated groundwater, emended description of the genus *Pelosinus*, and transfer of *Sporotalea propionica* to *Pelosinus propionicus* comb. nov. Int J Syst Evol Microbiol.

56. **Morris, J. J., R. E. Lenski, and E. R. Zinser.** 2012. The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss. MBio **3**.

57. **Mosher, J. J., T. J. Phelps, M. Podar, R. A. Hurt, Jr., J. H. Campbell, M. M. Drake, J. G. Moberly, C. W. Schadt, S. D. Brown, T. C. Hazen, A. P. Arkin, A. V. Palumbo, B. A. Faybishenko, and D. A. Elias.** 2012. Microbial community succession during

lactate amendment and electron acceptor limitation reveals a predominance of metal-reducing *Pelosinus* spp. Appl Environ Microbiol **78:**2082-91.

58. **Mulder, N., and R. Apweiler.** 2007. InterPro and InterProScan: tools for protein sequence classification and comparison. Methods Mol Biol **396:**59-70.

59. **Odom, J. M., and H. D. Peck, Jr.** 1984. Hydrogenase, electron-transfer proteins, and energy coupling in the sulfate-reducing bacteria *Desulfovibrio*. Annu Rev Microbiol **38:**551-92.

60. **Oslowski, D. M., J. H. Jung, D. H. Seo, C. S. Park, and J. F. Holden.** 2011. Production of hydrogen from alpha-1,4- and beta-1,4-linked saccharides by marine hyperthermophilic Archaea. Appl Environ Microbiol **77:**3169-73.

61. **Pankhania, I. P., A. N. Moosavi, and W. A. Hamilton.** 1986. Utilization of Cathodic Hydrogen by *Desulfovibrio vulgaris* (Hildenborough). Journal of General Microbiology **132:**3357-3365.

62. **Passerini, A., M. Lippi, and P. Frasconi.** 2011. MetalDetector v2.0: predicting the geometry of metal binding sites from protein sequence. Nucleic Acids Res **39:**W288-92.

63. **Pereira, P. M., Q. He, F. M. Valente, A. V. Xavier, J. Zhou, I. A. Pereira, and R. O. Louro.** 2008. Energy metabolism in *Desulfovibrio vulgaris* Hildenborough: insights from transcriptome analysis. Antonie Van Leeuwenhoek **93:**347-62.

64. **Poole, F. L., 2nd, B. A. Gerwe, R. C. Hopkins, G. J. Schut, M. V. Weinberg, F. E. Jenney, Jr., and M. W. Adams.** 2005. Defining genes in the genome of the hyperthermophilic archaeon *Pyrococcus furiosus*: implications for all microbial genomes. J Bacteriol **187:**7325-32.

65. **Porat, I., T. A. Vishnivetskaya, J. J. Mosher, C. C. Brandt, Z. K. Yang, S. C. Brooks, L. Liang, M. M. Drake, M. Podar, S. D. Brown, and A. V. Palumbo.** 2010. Characterization of archaeal community in contaminated and uncontaminated surface stream sediments. Microb Ecol **60:**784-95.

66. **Postgate, J. R.** 1951. On the nutrition of *Desulphovibrio desulphuricans*. J Gen Microbiol **5:**714-24.

67. **Postgate, J. R.** 1953. Presence of cytochrome in an obligate anaerobe. Biochem J **56:**xi-xii.

68. **Price, M. N., A. M. Deutschbauer, J. V. Kuehl, H. Liu, H. E. Witkowska, and A. P. Arkin.** 2010. Evidence-based annotation of transcripts and proteins in the sulfate-reducing bacterium *Desulfovibrio vulgaris* Hildenborough. J Bacteriol **193:**5716-27.

69. **Punta, M., and Y. Ofran.** 2008. The rough guide to in silico function prediction, or how to use sequence and structure information to predict protein function. PLoS Comput Biol **4:**e1000160.

70. **Qafoku, N. P., P. Evan Dresel, E. Ilton, J. P. McKinley, and C. T. Resch.** 2010. Chromium transport in an acidic waste contaminated subsurface medium: the role of reduction. Chemosphere **81:**1492-500.

71. **Ramos, A. R., K. L. Keller, J. D. Wall, and I. A. Pereira.** 2012. The Membrane QmoABC Complex Interacts Directly with the Dissimilatory Adenosine 5'-Phosphosulfate Reductase in Sulfate Reducing Bacteria. Front Microbiol **3:**137.

72. **Ray, A. E., J. R. Bargar, V. Sivaswamy, A. C. Dohnalkova, Y. Fujita, B. M. Peyton, and T. S. Magnuson.** 2011. Evidence for multiple modes of uranium immobilization by an anaerobic bacterium. Geochimica et Cosmochimica Acta **75:**2684-2695.

73. **Ray, A. E., S. A. Connon, P. P. Sheridan, J. Gilbreath, M. Shields, D. T. Newby, Y. Fujita, and T. S. Magnuson.** 2010. Intragenomic heterogeneity of the 16S rRNA gene in strain UFO1 caused by a 100-bp insertion in helix 6. FEMS Microbiol Ecol **72:**343-53.

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

75. **Sevcenco, A. M., M. W. Pinkse, E. Bol, G. C. Krijger, H. T. Wolterbeek, P. D. Verhaert, P. L. Hagedoorn, and W. R. Hagen.** 2009. The tungsten metallome of *Pyrococcus furiosus*. Metallomics **1:**395-402.

76. **Shelobolina, E. S., K. P. Nevin, J. D. Blakeney-Hayward, C. V. Johnsen, T. W. Plaia, P. Krader, T. Woodard, D. E. Holmes, C. G. Vanpraagh, and D. R. Lovley.** 2007. *Geobacter pickeringii* sp. nov., *Geobacter argillaceus* sp. nov. and *Pelosinus fermentans* gen. nov., sp. nov., isolated from subsurface kaolin lenses. Int J Syst Evol Microbiol **57:**126-35.

77. **Shirai, M., T. Taniguchi, and H. Kambara.** 2012. Emerging Applications of Single-Cell Diagnostics. Top Curr Chem.

78. **Singh, S., S. H. Kang, A. Mulchandani, and W. Chen.** 2008. Bioremediation: environmental clean-up through pathway engineering. Curr Opin Biotechnol **19:**437-44.

79. **Sodhi, J. S., K. Bryson, L. J. McGuffin, J. J. Ward, L. Wernisch, and D. T. Jones.** 2004. Predicting metal-binding site residues in low-resolution structural models. J Mol Biol **342:**307-20.

80. **Szpunar, J.** 2005. Advances in analytical methodology for bioinorganic speciation analysis: metallomics, metalloproteomics and heteroatom-tagged proteomics and metabolomics. Analyst **130:**442-65.

81. **Trauger, S. A., E. Kalisak, J. Kalisiak, H. Morita, M. V. Weinberg, A. L. Menon, F. L. Poole, 2nd, M. W. Adams, and G. Siuzdak.** 2008. Correlating the transcriptome, proteome, and metabolome in the environmental adaptation of a hyperthermophile. J Proteome Res **7:**1027-35.

82. **Treangen, T. J., and S. L. Salzberg.** 2012. Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat Rev Genet **13:**36-46.

83. **Vaccaro, B. J., A. L. Menon, W. A. Lancaster, and M. W. Adams.** 2012. Metallomics using Inductively Coupled Plasma Mass Spectrometry. Current Protocols in Chemical Biology **(in press)**.

84. **Vaudel, M., J. M. Burkhart, D. Breiter, R. P. Zahedi, A. Sickmann, and L. Martens.** 2012. A complex standard for protein identification, designed by evolution. J Proteome Res.

85. **Verhaart, M. R., A. A. Bielen, J. van der Oost, A. J. Stams, and S. W. Kengen.** 2010. Hydrogen production by hyperthermophilic and extremely thermophilic bacteria and archaea: mechanisms for reductant disposal. Environ Technol **31:**993-1003.

86. **Walker, C. B., Z. He, Z. K. Yang, J. A. Ringbauer, Jr., Q. He, J. Zhou, G. Voordouw, J. D. Wall, A. P. Arkin, T. C. Hazen, S. Stolyar, and D. A. Stahl.** 2009. The electron transfer system of syntrophically grown *Desulfovibrio vulgaris*. J Bacteriol **191:**5793-801.

87. **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.** 2005. Protein production and crystallization at SECSG -- an overview. J Struct Funct Genomics **6:**233-43.

88. **White, J. R., P. Escobar-Paramo, E. F. Mongodin, K. E. Nelson, and J. DiRuggiero.** 2008. Extensive genome rearrangements and multiple horizontal gene transfers in a population of *Pyrococcus* isolates from Vulcano Island, Italy. Appl Environ Microbiol **74:**6447-51.

89. **Yang, S. J., B. C. Min, Y. W. Kim, S. M. Jang, B. H. Lee, and K. H. Park.** 2007. Changes in the catalytic properties of *Pyrococcus furiosus* thermostable amylase by mutagenesis of the substrate binding sites. Appl Environ Microbiol **73:**5607-12.

90. **Yoon, S. H., D. J. Reiss, J. C. Bare, D. Tenenbaum, M. Pan, J. Slagel, R. L. Moritz, S. Lim, M. Hackett, A. L. Menon, M. W. Adams, A. Barnebey, S. M. Yannone, J. A. Leigh, and N. S. Baliga.** 2011. Parallel evolution of transcriptome architecture during genome reorganization. Genome Res **21:**1892-904.

91. **Zhang, Y., and V. N. Gladyshev.** 2011. Comparative genomics of trace element dependence in biology. J Biol Chem **286:**23623-9.

92. **Zillig, W., I. Holz, H.-P. Klenk, J. Trent, S. Wunderl, D. Janekovic, E. Imsel, and B. Haas.** 1987. *Pyrococcus woesei*, sp. nov., an ultra-thermophilic marine archaebacterium, representing a novel order, Thermococcales. Systematic and Applied Microbiology **9:**62-70.

93. **Zivanovic, Y., P. Lopez, H. Philippe, and P. Forterre.** 2002. *Pyrococcus* genome comparison evidences chromosome shuffling-driven evolution. Nucleic Acids Res **30:**1902-10.

Figure 1.1 Genome arrangements of the *P. furiosus* NCBI reference sequence and *P. furiosus* COM1. This diagram shows a Mauve genome alignment of the complete genome of the naturally competent *P. furiosus* strain COM1 with the reference sequence obtained in 2001. There is a high degree of overall synteny, with two major inversions, the AE (PF0001-PF0189, PF1603-PD2065) block and the C (PF0349-PF0388) block. IS elements flanking the C block are annotated in the NCBI reference, while recent transposition resulted in IS elements insertion into regions flanking the AE block. Modified from (9).

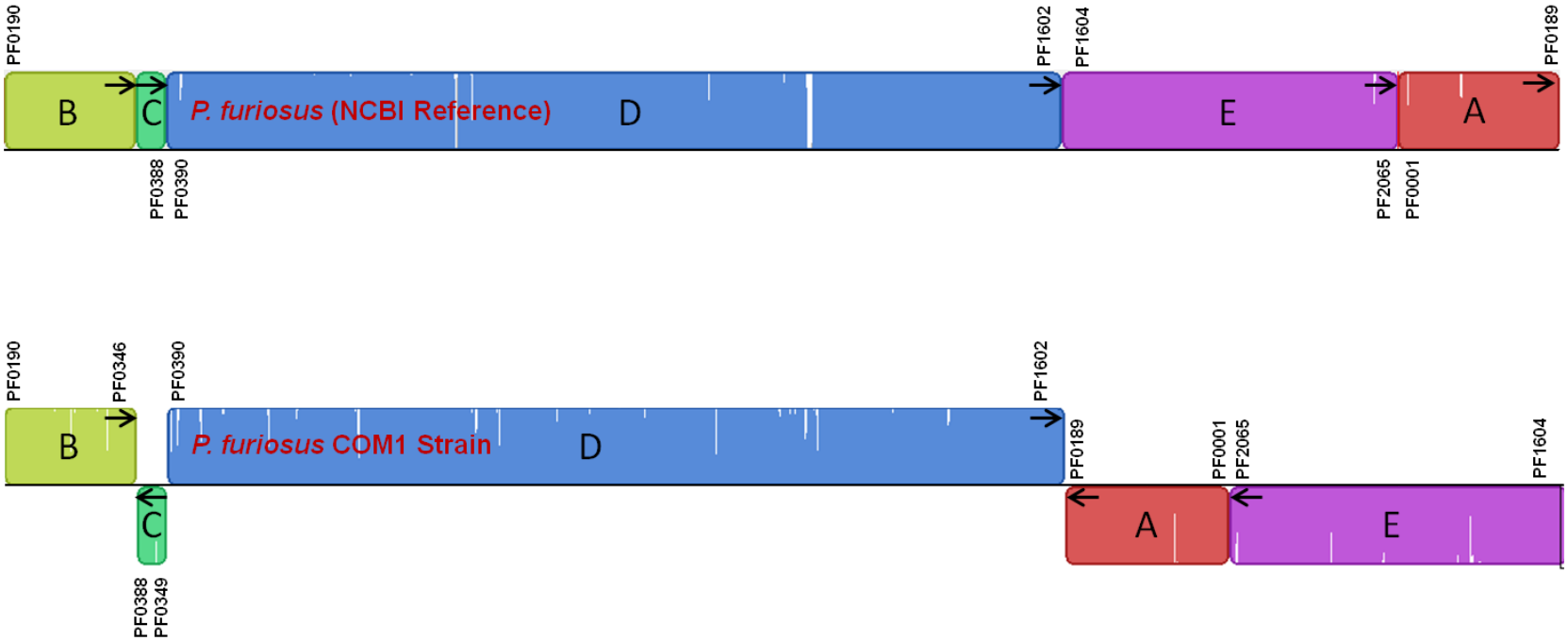
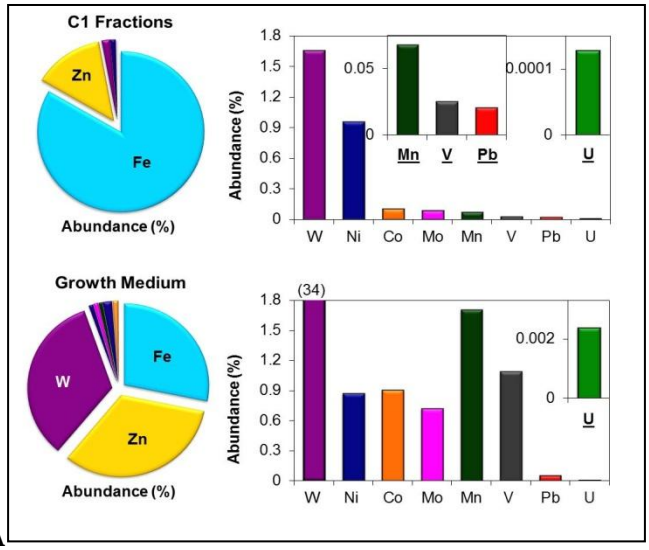
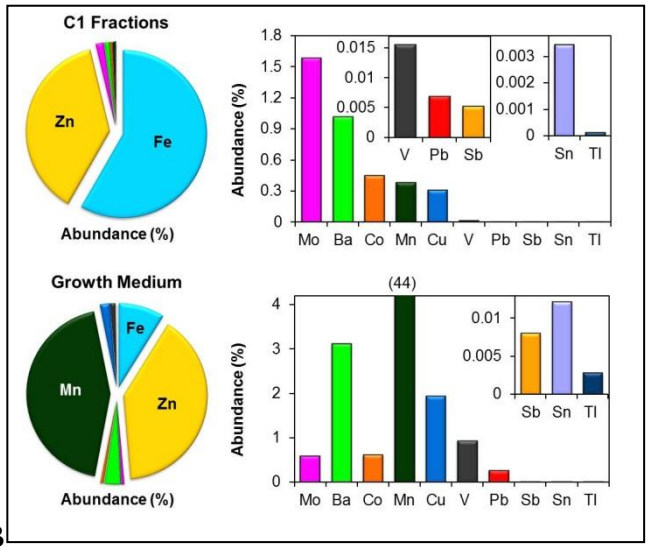

PF0190
B
C
P. furiosus (NCBI Reference)
D
PF1602
PF1604
E
A
PF0189
PF0388
PF0390
PF2065
PF0001
PF0190
PF0346
PF0390
B
P. furiosus COM1 Strain
D
PF1602
PF0189
PF0001
PF2065
PF1604
A
E
C
PF0388
PF0349

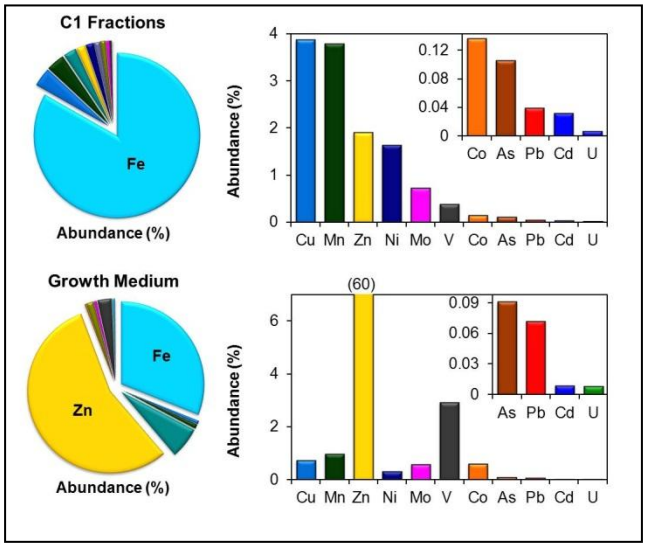
Figure 1.2. Metals present in the growth media and first level chromatography fractions (C1) of three organsisms, *Pyrococcus furiosus* (A), *Sulfolobus solfataricus* (B), and *Escherichia coli* (C). Significant differences were found in the relative amounts of various metals in the cytoplasm of three organisms. In addition, W was unique to *P. furiosus*, Cd and As were found only in *E. coli*, while *S. solfataricus* contained Sn,Sb, and Ba. Modified from(16).


C1 Fractions
Zn
Fe
Abundance (%)
Abundance (%)
1.8
1.5
1.2
0.9
0.6
0.3
0
0.05
0
0.0001
0
Mn V Pb
U
W Ni Co Mo Mn V Pb U
Growth Medium
W
Fe
Zn
Abundance (%)
(34)
0.002
0
U
W Ni Co Mo Mn V Pb U

A

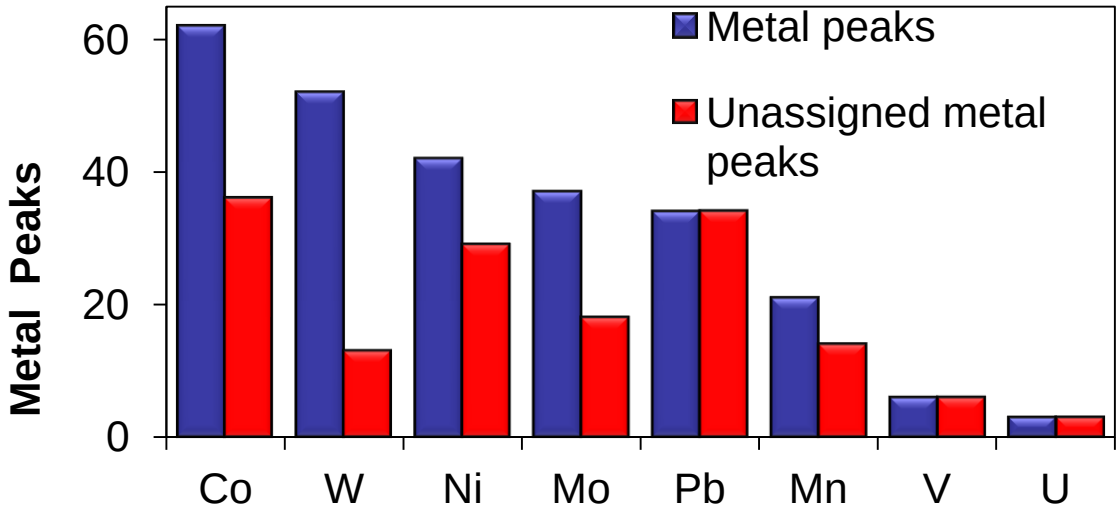

C1 Fractions
Zn
Fe
Abundance (%)
Abundance (%)
1.8
1.5
1.2
0.9
0.6
0.3
0
0.015
0.01
0.005
0
V Pb Sb
0.003
0.002
0.001
0
Sn Tl
Mo Ba Co Mn Cu V Pb Sb Sn Tl
Growth Medium
Mn
Fe
Zn
Abundance (%)
(44)
4
3
2
1
0
0.01
0.005
0
Sb Sn Tl
Mo Ba Co Mn Cu V Pb Sb Sn Tl

B


C1 Fractions
Fe
Abundance (%)
Abundance (%)
4
3
2
1
0
0.12
0.08
0.04
0
Co As Pb Cd U
Cu Mn Zn Ni Mo V Co As Pb Cd U
Growth Medium
Fe
Zn
Abundance (%)
(60)
6
4
2
0
0.09
0.06
0.03
0
As Pb Cd U
Cu Mn Zn Ni Mo V Co As Pb Cd U

C

Figure 1.3. Metal peaks obtained in the native fractionation of *P. furiosus* biomass. The proteins identified in each fraction comprising a metal peak were analyzed using IPM**.** If no protein was predicted to contain the observed metal, it was labeled "unassigned" and represents one or more novel metalloproteins. Modified from (16)


Metal Peaks
60
40
20
0
Co
W
Ni
Mo
Pb
Mn
V
U
Metal peaks
Unassigned metal peaks

CHAPTER 2

A COMPUTATIONAL FRAMEWORK FOR PROTEOME-WIDE PURSUIT AND PREDICTION OF METALLOPROTEINS USING ICP-MS AND MS/MS DATA[1]

[1] Lancaster*, W. A., J. L. Praissman*, F. L. Poole, 2nd, A. Cvetkovic, A. L. Menon, J. W. Scott, F. E. Jenney, Jr., M. P. Thorgersen, E. Kalisiak, J. V. Apon, S. A. Trauger, G. Siuzdak, J. A. Tainer, and M. W. Adams. (2011) BMC Bioinformatics 12**:**64. doi: 10.1186/1471-2105-12-64. Reproduced in accordance with BioMed Central's Open Access Charter.
*Contributed equally to this work

ABSTRACT

Background

Metal-containing proteins comprise a diverse and sizable category within the proteomes of organisms, ranging from proteins that use metals to catalyze reactions to proteins in which metals play key structural roles. Unfortunately, reliably predicting that a protein will contain a specific metal from its amino acid sequence is not currently possible. We recently developed a generally-applicable experimental technique for finding metalloproteins on a genome-wide scale. Applying this metal-directed protein purification approach (ICP-MS and MS/MS based) to the prototypical microbe *Pyrococcus furiosus* conclusively demonstrated the extent and diversity of the uncharacterized portion of microbial metalloproteomes since a majority of the observed metal peaks could not be assigned to known or predicted metalloproteins. However, even using this technique, it is not technically feasible to purify to homogeneity all metalloproteins in an organism. In order to address these limitations and complement the metal-directed protein purification, we developed a computational infrastructure and statistical methodology to aid in the pursuit and identification of novel metalloproteins.

Results

We demonstrate that our methodology enables predictions of metal-protein interactions using an experimental data set derived from a chromatography fractionation experiment in which 870 proteins and 10 metals were measured over 2,589 fractions. For each of the 10 metals, cobalt, iron, manganese, molybdenum, nickel, lead, tungsten, uranium, vanadium, and zinc, clusters of proteins frequently occurring in metal peaks (of a specific metal) within the fractionation space were defined. This resulted in predictions that there are from 5 undiscovered

vanadium- to 13 undiscovered cobalt-containing proteins in *Pyrococcus furiosus.* Molybdenum and nickel were chosen for additional assessment producing lists of genes predicted to encode metalloproteins or metalloprotein subunits, 22 for nickel including seven from known nickel-proteins, and 20 for molybdenum including two from known molybdo-proteins. The uncharacterized proteins are prime candidates for metal-based purification or recombinant approaches to validate these predictions.

Conclusions

We conclude that the largely uncharacterized extent of native metalloproteomes can be revealed through analysis of the co-occurrence of metals and proteins across a fractionation space. This can significantly impact our understanding of metallobiochemistry, disease mechanisms, and metal toxicity, with implications for bioremediation, medicine and other fields.

BACKGROUND

Metallomics is an emerging field that seeks to comprehensively characterize the role of metals in organisms (16). As with any new field, unique challenges have been encountered in terms of experimental methodologies and data analysis. The essential role metals play in biology has long been appreciated, but the complete metallome of any organism has yet to be characterized. It is estimated that around a third of all proteins in an organism require a metal partner (27). While it is possible to predict that certain proteins contain a metal (2), there are fundamental limitations to all current computational methods in comprehensively describing the metalloproteome of any organism. Many current methods rely on sequence motifs that in turn depend on the existence of a sufficiently sized set of previously annotated homologous proteins,

a problem further compounded by the diversity of metal-binding sites across organisms as well as within a single organism (7). It is also well known that heterologous protein expression can result in the production of proteins incorporating metals that are not natively incorporated (26). In addition, while certain proteins and protein families are known to bind a variety of metals and are annotated accordingly, many are annotated as binding a single metal based on limited evidence.

It was shown recently (5) that the set of metals known to interact with proteins in vivo is a significant underestimate of the true extent and diversity of the metalloproteome. The approach developed used metal-directed protein purification relying on inductively coupled plasma mass spectrometry (ICP-MS) and tandem mass spectrometry (MS/MS) and revealed that a prototypical microbe, *Pyrococcus furiosus*, takes up 21 of 53 metals measured in its growth medium, 18 of which are present in macromolecular complexes. These results are in stark contrast to the five metals that had previously been identified in proteins individually purified from the same organism. Further, of the 343 metal peaks found across the fractions from a second level of chromatography fractionation (for the 10 metals that were detected), almost half (158) contained no known or predicted metalloprotein corresponding to that particular metal (5). The purification of eight of these metal peaks resulted in the identification of novel metalloproteins, or proteins containing unexpected metal ions (5). Unfortunately, this method has two major limitations. Firstly, given the large number (158) of unassigned metal peaks and difficulty in purifying a single protein, it is impractical to purify a significant portion of these novel metalloproteins. Secondly, it is not technically feasible to natively purify proteins of very low abundance over several chromatographic steps.

Herein is described a computational infrastructure and analytical methodology developed to both aid in the pursuit of novel metalloproteins (15) as well as to predict which proteins observed via MS/MS during this fractionation are likely to be metalloproteins without requiring purification to homogeneity. This required the development of a database, an Online Analytical Processing (OLAP) cube and InterPro-Metal (IPM) automated metal domain identification methods (directly supporting the pursuit of novel metalloproteins), as well as a Global Metal Protein Association (GMPA) analysis (enabling the prediction of metal-protein associations without complete purification). Given the essential biological role of metals, the discovery of novel metalloproteins has a multitude of implications in a variety of fields (9, 16). Moreover, the computational infrastructure and methods described can be applied to any form of biomass of interest from tissues to microbes to identify potential metalloprotein targets for experimental characterization. Most importantly, this analysis allows the discovery of low abundance metalloproteins without radioisotope labeling, which have eluded other methods (21, 25) but which nonetheless may occupy key roles in essential biological pathways.

METHODS

Experimental design

The experimental data set utilized in this study is an expanded version of the data set described in (15). Briefly, native biomass of the hyperthermophilic archaeon *Pyrococcus furiosus* was fractionated anaerobically through multiple non-denaturing chromatography steps utilizing multiple column chemistries. The resulting 2,589 fractions were analyzed by ICP-MS to identify metals, and by MS/MS to identify proteins (primarily high-throughput MS/MS in this study). The MS/MS data were filtered such that the false discovery rate was less than 1%, as

described in (15) and only proteins identified by Mascot with two or more peptides were considered in the current bioinformatics based study. The use of non-denaturing native chromatography, ICP-MS and MS/MS captures the co-occurrence of metals and proteins in their native form, and enabled a metal-based purification strategy, in contrast to conventional enzyme assay guided protein purification (5). While this metal-based separation was used to purify a number of metalloproteins to homogeneity (5), the wealth of metal and protein data collected for proteins that were not explicitly targeted for direct purification provided an additional opportunity (applying data analysis techniques) to identify proteins that are likely to contain one or more metals in their native form.

Data infrastructure

A relational database was constructed using Microsoft SQL Server 2005 to store the data used in this study. The database consists of three principal modules: a procedural (fractionation) module, a metal data (ICP-MS) module, and a protein data (MS/MS) module. The fractionation module was designed to store the procedural information used in each of the separation steps carried out during the multi-level column fractionation (allowing reconstruction of the complete experimental pathway). These multi-level hierarchical relationships between fractions were queried using recursive common table expressions (CTE). The metal data module was designed to store both procedural data and replicate metal concentration data for each sample and metal analyzed using ICPMS. This module also stores the peak assignments determined by manual inspection. Finally, the protein module stores data for each peptide identified by MS/MS, its protein source and the corresponding ORF, and details related to the MS/MS run and Mascot search as imported from Mascot XML result files. All fractions and samples (from fractions)

collected were assigned unique IDs and labeled with 2D data matrix barcodes to facilitate sample tracking. This ensured the simple and reliable association of the data contained in all three principal modules within the context of the experimental hierarchy.

The relational database contains 2,589 records corresponding to fractions (the fractionation module) obtained from the chromatographic separation in which 1,026 proteins (the protein module) were identified by MS/MS. Of these 1,026 proteins (corresponding to 135,989 peptide hit records in the database), 870 were identified with 2 or more peptides (125,777 peptide hit records) and were used in statistical scoring of metal association (GMPA scores, see below). Each of these fractions has associated metal concentration data generated by ICP-MS analysis for multiple metals depending in part on what metals were relevant to the goals of a given separation step (up to 53 with 78,514 overall metal concentration records; the metal module). An Online Analytical Processing (OLAP) cube (Figure 1, middle right) was constructed on top of the relational database using Microsoft SQL Server 2005's Analysis Services. The cube enabled efficient preprocessing of significant amounts of aggregate data (e.g. sums, averages, etc.) and at the same time enabled convenient data slicing across multiple dimensions of experimental data in real-time (e.g. queries as to the number of proteins and metals detected in the same fraction) (12). OLAP is commonly used in the business analysis field but has not been widely exploited in scientific fields. However, OLAP is particularly well suited to the types of analyses presented herein given the hierarchical nature of the data set and the aggregate nature of queries utilized in this study to investigate experimental outcomes (e.g. are we purifying a protein – i.e. are metal/protein ratios improving?). A more detailed description of data infrastructure is given in the supplementary material (Additional Files 1, 2, 3, 4, 5). For the 870 proteins with GMPA scores, metal concentrations and numbers of peptides in

the column fractions containing those proteins (designated by ORF number) are available at http://enigma.bmb.uga.edu/impact.

Bioinformatic metalloprotein prediction – InterPro-Metal (IPM) automated metal domain analysis

The set of known metalloproteins that have been previously purified and characterized from *P. furiosus* by conventional chromatographic methods consists of only 23 proteins (encoded by 39 genes). Each contains one or more of Co, Fe, Ni, W and/or Zn atoms (5). Although the utility of bioinformatic predictions is limited, such predictions can be used to identify homologs of more extensively characterized metalloproteins and serve as a starting point for assigning proteins to the observed metal peaks. The Integrated Resource of Protein Domains and Functional Sites (InterPro) (10) was used to predict known metal associated domains encoded in the genome of *P. furiosus*. InterPro integrates multiple popular protein feature databases, and provides the Iprscan utility for searching protein sequence queries against these databases. The genome was searched using this utility and the resulting matches of proteins to InterPro entries were stored in a relational database (protein-InterPro data). The description of each InterPro entry, including name, abstract and publication list, is available in a downloadable XML file (ftp://ftp.ebi.ac.uk/pub/databases/interpro/interpro.xml.gz). This file was parsed and inserted into a corresponding relational database. A number of regular expression patterns relating to metal ions, metal cofactors and metal binding domains were used to search the text of each InterPro entry description to classify the entries which potentially involve specific metals (metal-InterPro data). Those metal-InterPro entries that had hits in the *P. furiosus* genome were evaluated manually for quality and assigned a subjective score. In some cases, keyword hits

were not deemed to be indicative of a potential association of the given domain with a given metal, for instance an abstract for a particular subfamily of an enzyme may include additional information on other subfamilies which use alternate metals. Such spurious hits were assigned a score of 0 while metals with evidence of association with the given domain were assigned a positive score. All hits with nonzero scores were considered as potentially metal associated domains in subsequent analyses. The protein-InterPro data and metal-InterPro data were joined to determine which *P. furiosus* proteins had associations with specific metals and this subsequently will be referred to as the InterPro-Metal (IPM) database or InterPro-Metal analysis. These domain-based predictions were incorporated into the relational database and OLAP cube (Figure 1) to aid in the identification of novel metalloproteins and proteins for which the metal prediction and observed metal associations differ.

Data driven metalloprotein prediction – Global Metal Protein Association (GMPA) analysis

The previously described infrastructure enables efficient querying of the data along the following dimensions: protein identity, metal motif prediction for identified proteins, total protein concentration and metal concentration across the entire experimental data set. This database enables an evaluation of the global association of metals with proteins across the entire observed space. The heterogeneity of the utilized data, with ICP-MS yielding quantitative metal concentrations and the MS/MS results indicating the presence or absence of proteins (with peptide counts providing only local semi-quantitative comparison), dictated the use of methods which are less reliant on quantitative agreement between these data sets. Metal peak fraction regions were defined and entered into the database manually based roughly on the presence of at

least two fractions in which the concentrations for a given metal (as indicated by two independent ICP-MS technical replicates) were substantially above the surrounding data for that metal. A hypergeometric distribution-based statistic that only considers the presence of proteins in these metal peaks instead of similarity of the shape or size of the metal and protein peaks (e.g. peptide counts of a given protein across a relatively contiguous set of fractions) was utilized. This statistic, the "Global Metal Protein Association score" (GMPA score, $G(p_i,m_j)$) is defined as:

$$G(p_i, m_j) = \sum_{n=\max\left(f_{p_i,m_j},\, f_{p_i} - \left(f - f_{m_j}\right)\right)}^{\min\left(f_{p_i},\ f_{m_j}\right)} \frac{\binom{f_{m_j}}{n}\binom{f - f_{m_j}}{f_{p_i} - n}}{\binom{f}{f_{p_i}}}$$

$$= phyper\left(f_{p_i,m_j} - 1, f_{m_j}, f - f_{m_j}, f_{p_i}, FALSE\right)$$

where f = the number of fractions in the data set, f_{m_j} = the number of fractions inside of peaks of metal *j* in the data set, f_{p_i} = the number of fractions in which protein *i* was observed, and f_{p_i,m_j} = the number of fractions in which protein *i* was observed that are also contained within peaks of metal *j*. This statistic gives the probability of at least f_{p_i,m_j} fractions out of the f_{p_i} fractions in which protein *i* is found occurring within peaks of metal *j* assuming that the f_{p_i} fractions were distributed randomly (with uniform probability) across the fraction space (Figure 2). The GMPA score was computed for all data sets we examined using the *phyper* function in R-2.10.1 (18) according to the formula above. The lower the score for a given protein/metal combination, the less likely it is that the protein and metal co-occur in the chromatographic fraction space by chance.

The frequency of observation plays a large role in the GMPA score, so an exponential significance curve through the GMPA score/occurrences space was introduced dividing it into a set of proteins with substantial evidence for metal association (lower GMPA scores, fewer fraction occurrences) and a set without as much evidence (higher GMPA scores, more fraction occurrences). This provides an initial filtering step for our current real-world data set (Figure 3). For metals significantly represented in the set of known metalloproteins (Fe, Ni, and W), the significance curve was generally set using these known references. For metals present in few or no known metalloprotein(s), the curve was extrapolated from the sets with sufficient known metalloproteins (Additional File 6). This extrapolation was based on the ratios between the exponent of an exponential regression curve calculated over all proteins through the GMPA score/occurrences space (Figure 3) and the significance curves chosen to capture all known metalloproteins that could reasonably be captured for Fe, Ni and W. Typically, an absolute GMPA score cut-off was employed after significance curve filtering to remove proteins that occurred in fewer than approximately 10 fractions from further analysis since it is unlikely that anything can be determined from our data with any confidence for such proteins.

To further organize the space of potential metalloproteins and to estimate the number of metalloproteins that could reasonably be expected from among the set of GMPA score filtered proteins for each metal, hierarchical clustering was applied. Hierarchical clustering organizes the elements of a set into a tree based on similarity (co-occurrence of proteins in fractions in this case). The resulting tree can facilitate the partitioning of the original set into sub-tree "clusters" capturing natural divisions in the data. This technique is particularly useful when the number of natural groups that might exist in the data is not known a priori. Ward's method of hierarchical clustering, which is variance-minimizing, along with the Euclidean metric and the Dynamic Tree

Cut package for R (13) were found to give useful self-contained clusters. The cutreeHybrid function of the Dynamic Tree Cut package was used to analyze the trees generated for each metal (Figure 4, Additional Files 7 and 8 – graphics produced in part using slightly modified versions of functions from (31)). The parameters passed to this function were tuned for Mo and Ni so that within the selected clusters there were clear "core" regions (composed of fractions with ≥50% of proteins in the cluster observed) while "core" region overlap was minimal between pairs of clusters. Low overlap between clusters in metal peak regions, with which core regions typically coincide, makes it likely that each cluster should contain at least one metalloprotein in order to explain the metal peak data – i.e. that the dimensionality reduction achieved in the clusterings is reasonable. Approximate minima required to cover all peaks were also calculated using the greedy algorithm approach to the set cover problem (4) and were found to be consistent with the results of clustering. The parameters found for Mo and Ni were then applied for all metals.

RESULTS AND DISCUSSION

The computational framework that was developed consisted of a database, an Online Analytical Processing (OLAP) cube, InterPro-Metal (IPM) automated metal domain identification and Global Metal Protein Association (GMPA) analysis. This complemented and enhanced our recent effort to elucidate the metalloproteome of *P. furiosus* and to identify novel metalloproteins (5). The GMPA analysis in particular was used to provide estimates of the numbers of metalloproteins that could be expected proteome-wide and a narrowed list of candidates (based on ICP-MS and MS/MS data) at various stages of the column chromatography fractionation, culminating in our predictions at the conclusion of the study that the numbers of

undiscovered metal containing proteins in *P. furiosus* range from approximately 5 for vanadium up to as many as 13 for cobalt. Validation of these predictions is provided by the overall success of the GMPA analysis in categorizing known metalloproteins from *P. furiosus*, the establishment of lower bounds on the numbers of proteins required to explain all metal peaks seen during the fractionation, the fact that the predictions lie within reasonable ranges in the context of literature (where up to a third of proteins are expected to contain or be involved with metals in various ways with the caveat that the majority are likely to be involved with Mg) (1, 27, 30) and considering the effect of dynamic association/adventitious metal binding (29).

In order to determine how much of the information contained in our data set remains uncaptured by the GMPA analysis, nickel (Ni) and molybdenum (Mo) were chosen for manual evaluation using the GMPA predictions as a starting point. Of the 870 proteins identified by MS/MS with two or more peptides, 153 and 119 were found to be significantly associated with Ni and Mo respectively, upon clustering yielding predictions of 13 and 10 total Ni- and Mo-proteins in the proteome of *P. furiosus*. The local semi-quantitative MS/MS peptide hit data for each of the proteins clustered was then manually compared to the local metal concentration data using our data explorer (Figure 1). This step excluded an additional 131 and 99 proteins producing top candidates lists of 22 and 20 proteins that are most likely to contain Ni and Mo respectively. These lists were then analyzed more extensively through bioinformatic analyses and literature searches. We will first describe the results obtained at the conclusion of our experimental study and then discuss the bioinformatics of the lists of predicted Ni- and Mo-proteins, concluding with the limitations inherent in this study.

Bioinformatic metalloprotein prediction – InterPro-Metal (IPM) automated metal domain analysis predictions

Of the 2,065 annotated opening reading frames (ORFs) in the RefSeq annotation of the *P. furiosus* genome (17), 376 were found to have matches to metal-associated InterPro entries. These included all of the 23 previously known metalloproteins (5). Of the 376, 221 were detected by MS/MS with 2 or more peptides, 43 of which had matches to multiple metals, either from matches to multiple InterPro entries, or to a single InterPro entry that lists potential associations with multiple metals. Consistent with expectation, the majority of the *P. furiosus* proteins with metal-related InterPro hits were predicted to be associated with Fe or Zn, with fewer Mn, Mo, Co, W and Ni predictions (Figure 5) (1, 7). There were no predictions of association with lead or uranium other than transport proteins whose InterPro descriptions may list many metal ions. The observation of a metal peak in a fraction in which no predicted metalloprotein was identified shows that one of the proteins identified is a completely novel metalloprotein, or one which uses a metal other than that expected by its annotation (5). An unexpectedly large number – 158 of the 353 metal peaks detected for the second level of column fractionation – were found in the data set.

Data driven metalloprotein prediction – Global Metal Protein Association (GMPA) analysis predictions

The 870 proteins identified by MS/MS with two or more peptides were assigned GMPA scores and partitioned into significant and insignificant regions as described in the methodology section. The number of proteins deemed significant ranged from 45 for V to 153 for Ni (Table 1). The proteins falling in the significant regions for each metal were then hierarchically

clustered by co-occurrence in fractions, with the number of clusters ranging from 5 for V to 16 for Co (Table 1). Each of these clusters is assumed to contain at least one metalloprotein giving rise to the observed metal peaks (see Methods). A total of 23 metalloproteins are known for *P. furiosus* from previous studies (5), and the coverage of the corresponding clusterings by these standards, together with the metalloproteins discovered by metal-directed purification (5), is summarized in Table 1. As a specific example of the data underlying each row of this table, Figure 4 shows the 13 clusters into which the 153 significant (by GMPA score) Ni targets fell. The two known Ni-containing enzymes, the soluble hydrogenases, lie in distinct clusters, as illustrated in Figure 4.

Nickel- and Molybdo-protein evaluation

Nickel and molybdenum were selected for further detailed analysis primarily using a data explorer developed in-house leveraging the speed of the OLAP cube (Figure 1). There are 12 genes that encode subunits of known Ni-containing proteins in *P. furiosus*. A total of seven of the 12 proteins encoded by these genes were detected by MS/MS analyses including five genes that encode the two multi-subunit Ni-containing soluble hydrogenases (I and II) of *P. furiosus*. These seven genes represent a set of positive controls for evaluation of the analysis. Prior to the metal-targeted comprehensive protein purification (5), *P. furiosus* was not known to have any Mo-proteins, so all proteins identified by that analysis represent novel Mo-proteins in this organism. For Ni, out of 870 proteins observed by MS/MS in at least one fraction of the fraction-space used for this analysis, 153 proteins were selected in the initial GMPA score significance curve screening. The parameters for the significance curve for Ni were chosen directly based on the known Ni-proteins such that the significance curve did not filter out any of

the subunits of the Ni-containing soluble hydrogenase I (PF0891-PF0894) and only filtered out one of the subunits of the Ni-containing soluble hydrogenase II (PF1330). This is essentially unavoidable given that PF1330 was found in relatively few (16) fractions. From the 153 proteins, 13 reasonably distinct clusters were defined after hierarchical clustering and each was manually evaluated (Table 2, Additional File 9). Seven of the clusters were found to contain proteins exhibiting local agreement of MS/MS data to the Ni ICP-MS data and 22 proteins or subunits of proteins were found in total (Table 2). Five of the 22 best candidates for Ni are in fact subunits of previously known Ni-protein complexes (Table 2) providing validation of this approach. An additional 2 of the 22 candidate Ni-proteins were very recently taken to purity or partial purity (PF0056 and PF0086) by the metal-directed purification and their assignment was confirmed (5). This leaves 15 potential novel Ni-containing proteins on which to carry out further experiments.

For Mo, out of 870 proteins observed by MS/MS in at least one fraction of the fraction-space used for this analysis, 119 proteins were selected in the initial GMPA score significance curve screening. The significance curve for Mo was selected (as described in Methods) by extrapolation based on the significance curves chosen for metals with known metalloproteins. From the 119 proteins, 10 clusters were defined after hierarchical clustering and each cluster was manually evaluated. Six of the clusters were found to contain proteins exhibiting local agreement of MS/MS data to the Mo ICP-MS data and 12 such proteins were found in total (Table 3, Additional File 9). A novel Mo-protein (PF1587) was purified by the metal-based chromatography method from which the data set was derived (5) and this is identified by the GMPA analysis as a likely Mo-protein. This nicely demonstrates the effectiveness of the method given sufficient observation. Finally, comparing the results of the IPM analysis to the manually evaluated GMPA analysis derived results, among the 22 manually evaluated targets deemed to

be likely Ni-proteins, only 3 had a Ni IPM match with 10 additional proteins having Co, Fe, Mn and Zn matches. Of the 20 manually evaluated targets deemed to be likely Mo-proteins, only 1 had a Mo IPM match.

Bioinformatic analyses of predicted Ni- and Mo-proteins

As discussed above, 7 of the 22 genes listed in Table 2 encode proteins or subunits of proteins which have been shown to contain Ni ions in *P. furiosus*. This includes subunits of soluble hydrogenase I (SHI) and soluble hydrogenase 2 (SHII) grouped in clusters 7 and 8 respectively. In addition, PF0615 in cluster 13 is annotated as a hypA protein, which is implicated in Ni insertion in the hydrogenases. The structure of a hypA homolog from *Thermococcus kodakaraensis* has been solved and its Ni-binding site described (28). This demonstrates that cluster 13 has at least two Ni-binding proteins that frequently co-occur in the fractionation space. Of the five ORFs in Table 2 with homologs whose crystal structures have been solved bound to metals other than Ni, three are now known to bind Ni (PF0056 and PF0086) in *P. furiosus* or are known to have a Ni-binding site (PF0615). In particular PF0086 has been shown to bind Ni (5), but its homolog from the closely related *Pyrococcus horikoshii* (PDB 2E18) was expressed recombinantly and crystallized in a Zn-bound form. This illustrates the flexibility of metal binding domains (6), and their ability to bind biologically incorrect metals when expressed heterologously (11). The two remaining ORFs with non-Ni homolog crystal structures are PF1664, which contains the cysteines that bind Zn (Cd in the crystal) in its homolog (23) and may be involved in binding Ni in *P. furiosus*, and PF2038 with a homolog that binds Mg-ATP. The only protein listed in Table 2 that is likely not to contain Ni is PF1861. This was previously purified from *P. furiosus* biomass and contained Co and Zn but not Ni (24). This

leaves 13 proteins that are predicted to contain Ni. These proteins have no known or conjecturable Ni associations based on their sequences and are assumed to predominantly contain a set of undiscovered Ni-binding sites. Finally, it is worth pointing out that PF0056 is one of five ORFs (PF0144, PF1987, PF0056, PF0138 and PF1500) annotated as either conserved hypothetical or with only domain/motif matches and has now been shown to bind Ni [6].

In contrast to the case of Ni, the pool of known molybdo-proteins in *P. furiosus* is small and far less can be ascertained bioinformatically. In particular, the role of Mo-proteins in *P. furiosus* is unclear, with only two such metalloproteins having recently been identified (5). Consequently, only two of 20 proteins in Table 3 have either been shown to bind Mo (PF1587) or have an IPM hit for Mo (PF0187). The other recently identified Mo-protein that was purified from *P. furiosus* (PF1972) was observed in only 17 chromatography fractions and was rejected by the GMPA analysis, which depends on sufficient levels of occurrence in the data set to establish significance of metal-protein association. On the positive side, many of the uncharacterized proteins contain residues that could be involved in Mo-binding (e.g. 14 of 20 contain at least one cysteine residue as is often involved in Mo-pterin binding) (3, 8, 19), but given the extent and complexity of typical molybdopterin binding interactions and biochemistry (20) and the lack of knowledge of Mo-binding in organisms closely related to *P. furiosus*, we have not looked into this aspect further. On the negative side, DNA polymerase (Mo cluster 9, Table 3), which has been well studied in many different organisms and is not known to bind or utilize Mo (although it is not clear if this has been directly considered previously) was picked as a top candidate Mo-protein. This illustrates that some of the targets that appear to reliably co-occur with a metal may be coincidental, or the result of interaction natively with additional proteins that are not strictly required for their primary function. Interestingly, four of the 20

predicted Mo-protein candidates have annotations that include “domain of unknown function.” The confirmation of Mo-binding, which has already occurred for PF1587 by metal-directed purification, should provide an improved foundation for functionally characterizing these conserved domains which so far have been elusive (5).

KNOWN LIMITATIONS

We initially attempted to use standard correlation-based statistical techniques such as principal component analysis (PCA) and canonical correlation analysis (CCA) to determine associations between metals and proteins based on the experimental data that were available (5). However, these efforts were hindered by the relatively non-quantitative nature of the MS/MS data available (lacking even spectral count information). Consequently, the GMPA analysis method was developed which is less reliant on quantitative agreement between these data sets. Simulated data sets demonstrated the effectiveness of GMPA scores alone given adequate separation regardless of the amount of noise observed in the peptide counts, but it was discovered that the metal-based fractionation typically did not produce a comprehensive enough data set containing an appropriate degree of overall separation. For example, the experimental data set is most consistent and comprehensive at the second column level (termed C2 in (15)) and separation is still relatively incomplete at this level. Consequently significance cut-off curves and clustering were employed completing the overall GMPA analysis methodology. It is likely that the predictive power of the methodology could be greatly improved by utilizing a data set with a more comprehensive fractionation across all levels, through the use of more quantitative MS/MS techniques (14, 22, 32) and more powerful statistical techniques (PCA/CCA) that could then be applied more easily. This methodology could also potentially be

carried out in a more automated fashion on an analytical scale to provide a rapid determination of the metalloproteins of any organism.

CONCLUSIONS

We have presented a computational methodology that can uncover probable metal-containing proteins using data from a non-comprehensive native fractionation coupled with metal and protein measurement using ICP-MS and MS/MS. This methodology has identified a number of candidate novel metalloproteins that are targets for future experimental verification. Application of the method to simulated data sets indicates that additional predictive accuracy could be achieved through the use of a more comprehensive fractionation. Our results for each of the 10 metals examined in this study underscore the unexplored complexity of metalloproteomes and have broad implications for protein structure and function as well as metal toxicity.

ABBREVIATIONS

The abbreviations used are: Co, cobalt; Fe, iron; Mn, manganese; Mo, molybdenum; Ni, nickel; Pb, lead; U, uranium; V, vanadium; W, tungsten; Zn, zinc; HT-MS/MS: high-throughput tandem mass spectrometry; ICP-MS: inductively coupled plasma mass spectrometry; ID: identifier; MS/MS: tandem mass spectrometry; OLAP: online analytical processing; ORF: open reading frame.

AUTHORS' CONTRIBUTIONS

W.A.L. developed the bioinformatic metalloprotein prediction approach. J.L.P. developed the OLAP infrastructure, data explorer software and Global Metal Protein Association

analysis approach and manually evaluated the Mo and Ni clusters. W.A.L. and J.L.P. performed and interpreted the results of bioinformatics analyses of manually evaluated candidates. F.L.P., W.A.L. and J.L.P. developed the database infrastructure. J.L.P., F.L.P. and W.A.L. developed the ENIGMA IMPACT website. F.L.P. and J.L.P. wrote data loaders and loaded MS/MS and chromatographic fractionation data into the database. F.L.P. and J.L.P. wrote ICP-MS data loading software and A.C. loaded all ICP-MS data and defined all metal peaks manually using this software. A.C. performed ICP-MS analyses. S.A.T., E.K., J.V.A. and G.S. performed HT-MS/MS analyses. A.C., A.L.M., M.P.T. and J.W.S. grew and fractionated *P. furiosus*. A.C., A.L.M., F.E.J., F.L.P., J.W.S., M.P.T. and J.A.T. and M.W.W.A. developed the original metal-directed purification. J.L.P, M.W.W.A. and W.A.L. wrote the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGEMENTS

This work conducted by ENIGMA was supported by the Office of Science, Office of Biological and Environmental Research, of the U. S. Department of Energy under Contract No. DE-AC02-05CH11231.

REFERENCES

1. **Andreini, C., I. Bertini, G. Cavallaro, G. L. Holliday, and J. M. Thornton.** 2008. Metal ions in biological catalysis: from enzyme databases to general principles. Journal Of Biological Inorganic Chemistry **13:**1205-1218.

2. **Andreini, C., I. Bertini, and A. Rosato.** 2009. Metalloproteomes: A Bioinformatic Approach. Acc Chem Res **42:**1471-1479.

3. **Brondino, C. D., M. J. Romao, I. Moura, and J. J. Moura.** 2006. Molybdenum and tungsten enzymes: the xanthine oxidase family. Curr Opin Chem Biol **10:**109-14.

4. **Chvatal, V.** 1979. A Greedy Heuristic for the Set-Covering Problem. MATHEMATICS OF OPERATIONS RESEARCH **4:**233-235.

5. **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.** 2010. Microbial metalloproteomes are largely uncharacterized. Nature **466:**779-82.

6. **Dudev, T., and C. Lim.** 2008. Metal Binding Affinity and Selectivity in Metalloproteins: Insights from Computational Studies. Annual Review of Biophysics **37:**97-116.

7. **Dupont, C. L., A. Butcher, R. E. Valas, P. E. Bourne, and G. Caetano-Anolles.** 2010. History of biological metal utilization inferred through phylogenomic analysis of protein structures. Proc Natl Acad Sci U S A **107:**10567-72.

8. **Gladyshev, V. N., S. V. Khangulov, M. J. Axley, and T. C. Stadtman.** 1994. Coordination of selenium to molybdenum in formate dehydrogenase H from Escherichia coli. Proc Natl Acad Sci U S A **91:**7708-11.

9. **Haferburg, G., and E. Kothe.** 2010. Metallomics: lessons for metalliferous soil remediation. Appl Microbiol Biotechnol **87:**1271-80.

10. **Hunter, S., R. Apweiler, T. K. Attwood, A. Bairoch, A. Bateman, D. Binns, P. Bork, U. Das, L. Daugherty, L. Duquenne, R. D. Finn, J. Gough, D. Haft, N. Hulo, D. Kahn, E. Kelly, A. Laugraud, I. Letunic, D. Lonsdale, R. Lopez, M. Madera, J. Maslen, C. McAnulla, J. McDowall, J. Mistry, A. Mitchell, N. Mulder, D. Natale, C. Orengo, A. F. Quinn, J. D. Selengut, C. J. Sigrist, M. Thimma, P. D. Thomas, F.**

Valentin, D. Wilson, C. H. Wu, and C. Yeats. 2009. InterPro: the integrative protein signature database. Nucleic Acids Res **37:**D211-5.

11. **Jenney, F. E., Jr., and M. W. Adams.** 2001. Rubredoxin from Pyrococcus furiosus. Methods Enzymol **334:**45-55.

12. **Kehl, C., A. M. Simms, R. D. Toofanny, and V. Daggett.** 2008. Dynameomics: a multi-dimensional analysis-optimized database for dynamic protein data. Protein Eng Des Sel **21:**379-86.

13. **Langfelder, P., B. Zhang, and S. Horvath.** 2008. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics **24:**719-20.

14. **Lu, P., C. Vogel, R. Wang, X. Yao, and E. M. Marcotte.** 2007. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. Nat Biotechnol **25:**117-24.

15. **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.** 2009. Novel multiprotein complexes identified in the hyperthermophilic archaeon *Pyrococcus furiosus* by non-denaturing fractionation of the native proteome. Mol Cell Proteomics **8:**735-51.

16. **Mounicou, S., J. Szpunar, and R. Lobinski.** 2009. Metallomics: the concept and methodology. Chemical Society Reviews **38:**1119-1138.

17. **Pruitt, K. D., T. Tatusova, and D. R. Maglott.** 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res **35:**D61-5.

18. **R Development Core Team** 2005, posting date. R : A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. [Online.]

19. **Romao, M. J., J. Knablein, R. Huber, and J. J. Moura.** 1997. Structure and function of molybdopterin containing enzymes. Prog Biophys Mol Biol **68:**121-44.

20. **Schwarz, G., R. R. Mendel, and M. W. Ribbe.** 2009. Molybdenum cofactors, enzymes and pathways. Nature **460:**839-47.

21. **Sevcenco, A. M., G. C. Krijger, M. W. Pinkse, P. D. Verhaert, W. R. Hagen, and P. L. Hagedoorn.** 2009. Development of a generic approach to native metalloproteomics: application to the quantitative identification of soluble copper proteins in *Escherichia coli*. J Biol Inorg Chem **14:**631-40.

22. **Silva, J. C., M. V. Gorenstein, G. Z. Li, J. P. Vissers, and S. J. Geromanos.** 2006. Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. Mol Cell Proteomics **5:**144-56.

23. **Sivaraman, J., R. S. Myers, L. Boju, T. Sulea, M. Cygler, V. Jo Davisson, and J. D. Schrag.** 2005. Crystal structure of Methanobacterium thermoautotrophicum phosphoribosyl-AMP cyclohydrolase HisI. Biochemistry **44:**10071-80.

24. **Story, S. V., C. Shah, F. E. Jenney, Jr., and M. W. Adams.** 2005. Characterization of a novel zinc-containing, lysine-specific aminopeptidase from the hyperthermophilic archaeon Pyrococcus furiosus. J Bacteriol **187:**2077-83.

25. **Tottey, S., K. J. Waldron, S. J. Firbank, B. Reale, C. Bessant, K. Sato, T. R. Cheek, J. Gray, M. J. Banfield, C. Dennison, and N. J. Robinson.** 2008. Protein-folding location can regulate manganese-binding versus copper- or zinc-binding. Nature **455:**1138-42.

26. **Waldron, K. J., and N. J. Robinson.** 2009. How do bacterial cells ensure that metalloproteins get the correct metal? Nat Rev Microbiol **7:**25-35.

27. **Waldron, K. J., J. C. Rutherford, D. Ford, and N. J. Robinson.** 2009. Metalloproteins and metal sensing. Nature **460:**823-30.

28. **Watanabe, S., T. Arai, R. Matsumi, H. Atomi, T. Imanaka, and K. Miki.** 2009. Crystal structure of HypA, a nickel-binding metallochaperone for [NiFe] hydrogenase maturation. J Mol Biol **394:**448-59.

29. **Xiao, Z., and A. G. Wedd.** 2010. The challenges of determining metal-protein affinities. Nat Prod Rep **27:**768-89.

30. **Yamaguchi, A., K. Iida, N. Matsui, S. Tomoda, K. Yura, and M. Go.** 2004. Het-PDB Navi.: a database for protein-small molecule interactions. J Biochem **135:**79-84.

31. **Zhang, B., and S. Horvath.** 2005. A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol **4:**Article17.

32. **Zhu, W., J. W. Smith, and C. M. Huang.** 2010. Mass spectrometry-based label-free quantitative proteomics. J Biomed Biotechnol **2010:**840518.

Table 2.1 – GMPA Analysis Clustering and Coverage of Known Metalloproteins

Metal	Known metalloprotein subunits			Proteins Clustered	Clusters	
	Total	Observed	Met GMPA significance criterion		Total	With known metalloproteins
Co	5	5	3	139	16	3
Fe	35(20)	35	19	148	15	8
Mn	0	0	0	73	7	0
Mo	2	2	1	119	10	1
Ni	12(5)	12	9	153	13	4
Pb	0	0	0	90	9	0
U	0	0	0	76	7	0
V	0	0	0	45	5	0
W	5	5	4	136	11	4
Zn	5	5	1	116	9	1

Some proteins consist of multiple subunits, the number of holoenzymes are given in parentheses. Some proteins contain multiple metals and factor into multiple rows. For full clusters, refer to supplementary material.

Table 2.2 - Manually Evaluated Nickel Protein Candidates

Cluster number	**ORF**	**Annotation**	**Crystal structure homolog**	**Metal In structure**	**IPM prediction**
2	PF0144	Aldolase-type TIM barrel			Fe
2	PF1881	Alba archaeal DNA/RNA-binding protein	2Z7C		
3	PF0038	Beta-lactamase-like glyoxalase II family member			Zn
4	PF1916	Glycosyl transferase, family 2			
4	PF1987	Conserved hypothetical protein			
5	***PF0056***	***Carbohydrate binding protein***	***1VJ2***	***Mn***	***Mn***
5	PF0138	Uncharacterized rubrerythrin domain protein	2FZF		Fe
6	PF1664	Phosphoribosyl-AMP cyclohydrolase	1ZPS	Cd	Zn
6	PF2038	Adenosylcobalamin biosynthesis	1G5T	Mg	Co
7	**PF1329**	**Hydrogenase II beta**			**Fe**

7	**PF1331**	**Hydrogenase II delta**			**Fe**
7	**PF1332**	**Hydrogenase II alpha**			**Fe,Ni**
8	**PF0891**	**Hydrogenase I beta**			**Fe**
8	**PF0894**	**Hydrogenase I alpha**			**Fe,Ni**
8	PF1500	PRC-barrel-like			
8	PF1529	Pyroxidine biosynthesis protein	2YZR		
12	PF1401	Peptidyl-prolyl cis-trans isomerase			
13	***PF0086***	***Alanyl-tRNA synthetase, class IIc***	***2E1B***	***Zn***	
13	PF0615	Hydrogenase expression/formation protein A	3A43	Zn	Ni
13	PF1272	LamB/YcsF	1V6T		
13	PF1684	Acetylglutamate kinase	2EGX		
13	PF1861	Lysyl aminopeptidase	2PE3		Zn

Cluster number refers to hierarchical clustering with dynamic hybrid partitioning, see Figure 4 for explanation and supplementary tables 1 – 8 in Additional File 9 for complete cluster tables. Crystal structures obtained from PDB with sequence similarity >50%.

ORF numbers in bold have been previously characterized in *P. furiosus*, those in italics were characterized by metal-targeted purification.

Table 2.3 - Manually Evaluated Molybdenum Protein Candidates

Cluster	ORF	Annotation	Crystal structure homolog	Metal in structure	IPM Metal
1	PF0009	ThiF family protein	1JWB	Zn	
1	PF0187	Putative cofactor synthesis protein			Fe,Mo,W
1	PF0668	YjgF-like protein	2DYY		
1	PF1718	Wyosine base formation, Radical SAM	2YX0		Fe
1	PF1766	Cell division transporter FtsY	3DMD		
2	PF1956	Fructose-1,6-bisphosphate aldolase class I	1OJX		
3	PF1828	Protein of unknown function DUF1621			
3	PF1886	Carbohydrate/purine kinase			
4	PF0098	NAD+ synthase	2E18		
4	PF0236	Phosphoribosyl pyrophosphokinase	1U9Y		Mg
4	PF1401	Peptidyl-prolyl cis-trans isomerase	1IX5		

4	PF1675	Asp/Glu/hydantoin racemase	2ZSK		
4	PF1731	Signal recognition particle 54	3DM5		
5	PF1538	Amidohydrolase 1	1P1M	Ni	
7	PF0523	Protein of unknown function DUF509	1ZD0	Mg	
7	PF1222	Protein of unknown function DUF217			
8	***PF1587***	***Protein of unknown function DUF89***	***2G8L***		
9	PF0212	DNA polymerase, family B	2JGU	Mn	
9	PF0306	Translation factor, SUA5 type			
9	PF0463	Phosphoglycolate phosphatase			

Cluster number refers to hierarchical clustering with dynamic hybrid partitioning, see Figure 4 for explanation and supplementary tables 1 – 8 in Additional File 9 for complete cluster tables. Crystal structures obtained from PDB with sequence similarity >50%. ORF numbers in bold have been previously characterized in *P. furiosus*, those in italics were characterized by metal-targeted purification.

Figure 2.1. Experimental and computational overview. Below the dashed line is a schematic overview of the framework: all of the tools, databases and methods developed and utilized in this effort. Subsequent figures focus on the methods and calculations in the bioinformatics category.

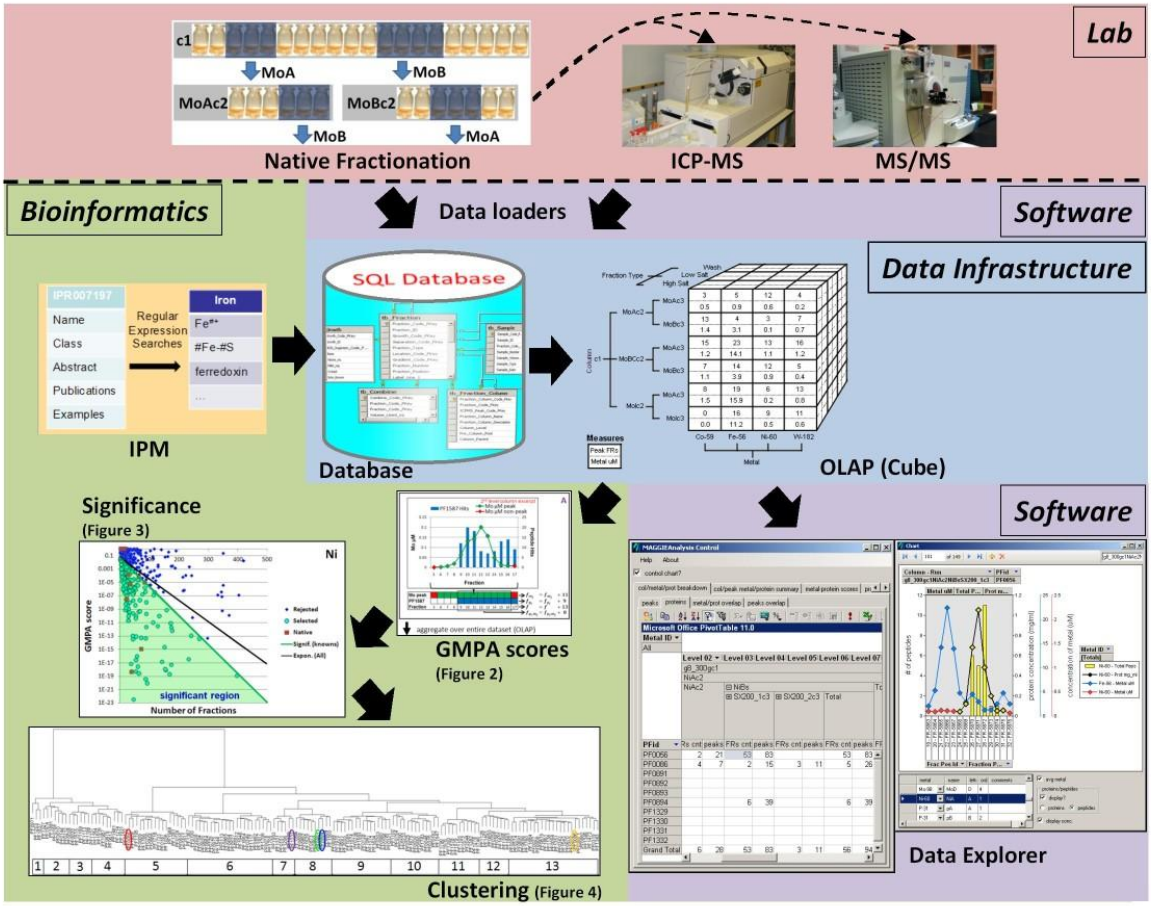
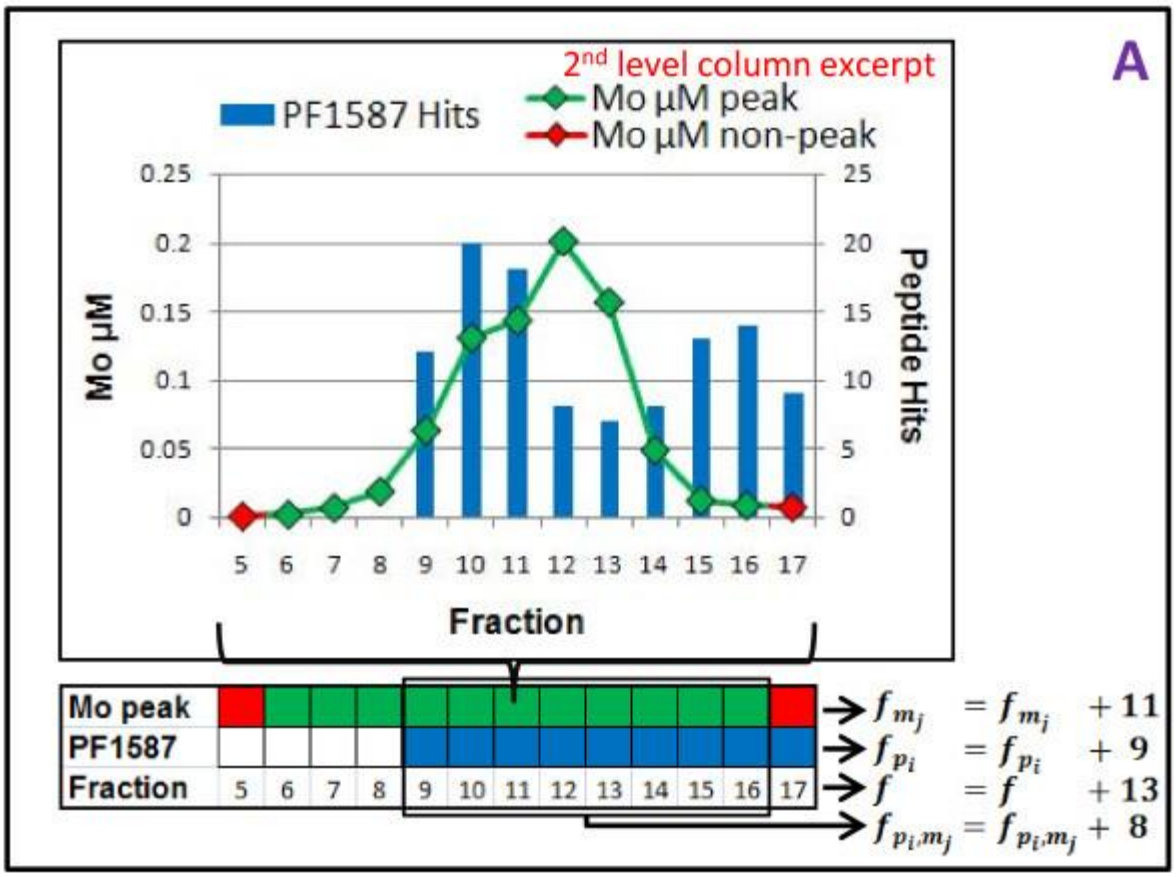

Lab
c1
MoA
MoB
MoAc2
MoBc2
MoB
MoA
Native Fractionation
ICP-MS
MS/MS
Bioinformatics
Data loaders
Software
Data Infrastructure
IPR007197
Name
Class
Abstract
Publications
Examples
Regular Expression Searches
Iron
Fe**
#Fe-#S
ferredoxin
IPM
SQL Database
Database
OLAP (Cube)
Significance
(Figure 3)
Ni
GMPA score
Number of Fractions
significant region
GMPA scores
(Figure 2)
Clustering (Figure 4)
Software
Data Explorer

Figure 2.2. GMPA score (Global Metal Protein Association) Calculation. A, B and C illustrate the calculation with data for p_i = PF1587 and m_j = Mo, arrows represent generic steps in the calculation. A) Peptide counts for each protein (per fraction) are reduced to Boolean values (present/not present-shown as blue/white cells respectively). Whether a fraction is part of a metal peak or not is already a Boolean value (green/red cells). B) The "present/not present" values are counted across all fractions in the data set. C) The GMPA score is calculated from these values using the hypergeometric distribution and is roughly a p-value: how likely is a given protein to have been seen in metal peak fractions as many times as it was (or more) assuming an equal likelihood for the protein to have been observed in any fraction and given the number of metal peak fractions, protein fractions, and total fractions.



aggregate over entire dataset (OLAP)

PF1587 Mo peak FRs	(f_{p_i,m_j})	187
PF1587 FRs	(f_{p_i})	235
Total Mo peak FRs	(f_{m_j})	1,030
Total FRs	(f)	2,589

B

$$phyper\left(f_{p_i,m_j} - 1, f_{m_j}, f - f_{m_j}, f_{p_i}, FALSE\right)$$

C

$$phyper(186, 1030, 1559, 235, FALSE) = 5.42E-39$$

Figure 2.3. Protein selection using GMPA score significance curve criterion. Scatterplots (per metal) of the GMPA scores versus number of fraction occurrences for all proteins. The green/orange lines are exponential "significance" curves (plotted on a logarithmic scale) and proteins below are considered significant and selected for clustering (green regions and cyan points). The Ni significance curve (green) was based on occurrences of known Ni-proteins (red points) and the Mo significance curve (orange) was extrapolated from the relationships between the significance curves and exponential regression curves (through all points-the black lines) on average across all metals with known metalloproteins. Typically, this step removes an additional 20-35 proteins beyond what would be removed using the regression curves themselves.

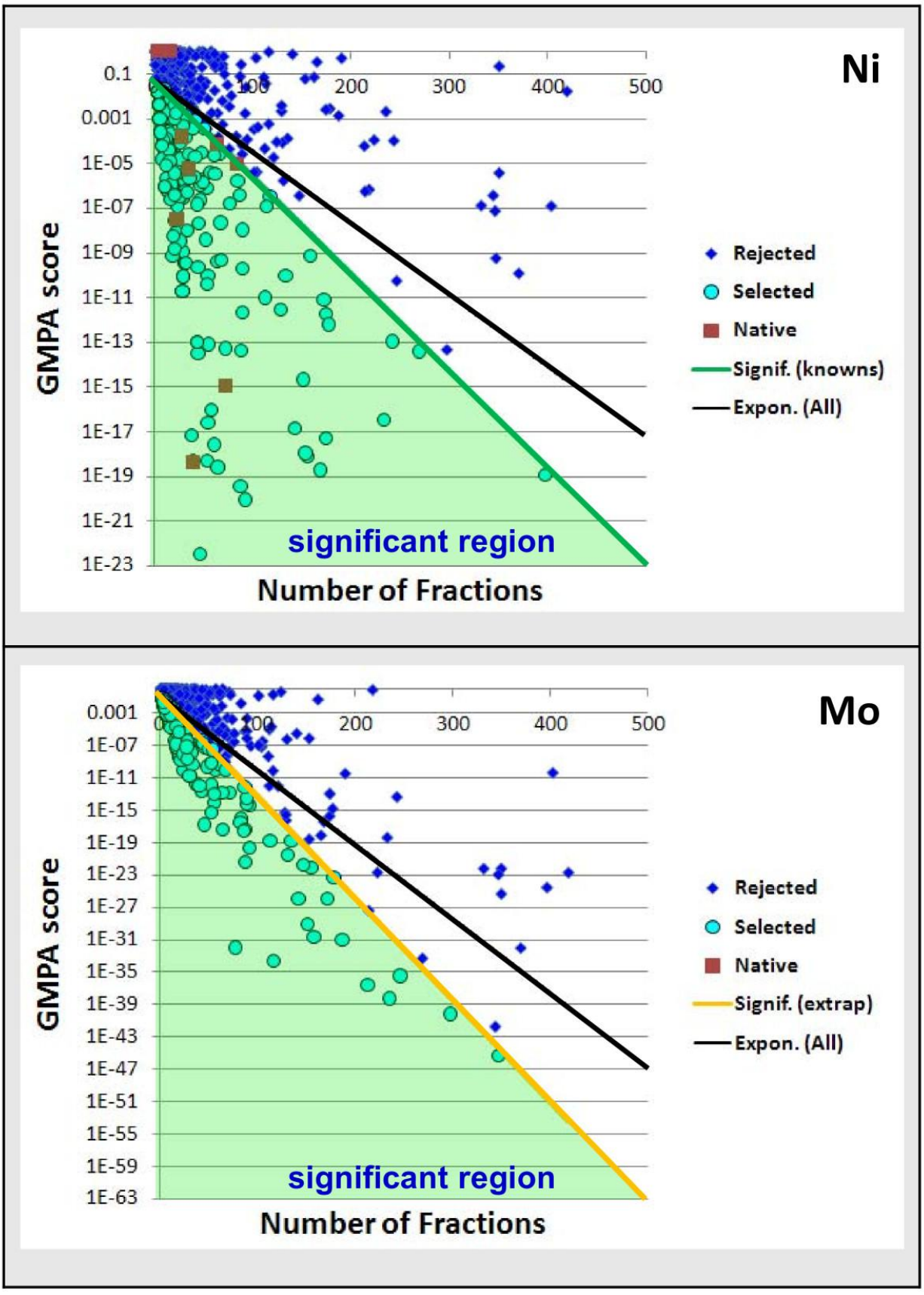

Ni
GMPA score
0.1
0.001
1E-05
1E-07
1E-09
1E-11
1E-13
1E-15
1E-17
1E-19
1E-21
1E-23
100
200
300
400
500
significant region
Number of Fractions
Rejected
Selected
Native
Signif. (knowns)
Expon. (All)
Mo
GMPA score
0.001
1E-07
1E-11
1E-15
1E-19
1E-23
1E-27
1E-31
1E-35
1E-39
1E-43
1E-47
1E-51
1E-55
1E-59
1E-63
100
200
300
400
500
significant region
Number of Fractions
Rejected
Selected
Native
Signif. (extrap)
Expon. (All)

Figure 2.4. Hierarchical clustering of GMPA score criterion significant nickel associated proteins. Potential nickel proteins clustered based on co-occurrence in fractions using Ward's method. The grid at top contains excerpts from the data set selected to give a rough sense of protein co-occurrence within fractions and its effect on the resulting clustering. The numbered boxes at the bottom indicate the partitioning of the overall clustering into self-contained clusters as determined by cutreeHybrid (indicated by colors in supplementary material clusterings). Clusters containing known Ni-proteins are highlighted at the bottom illustrating the clustering together of subunits of known nickel-proteins and clustering apart of distinct Ni-proteins (SHI: soluble hydrogenase I PF0891-PF0894, SHII: soluble hydrogenase II PF1329-PF1332).

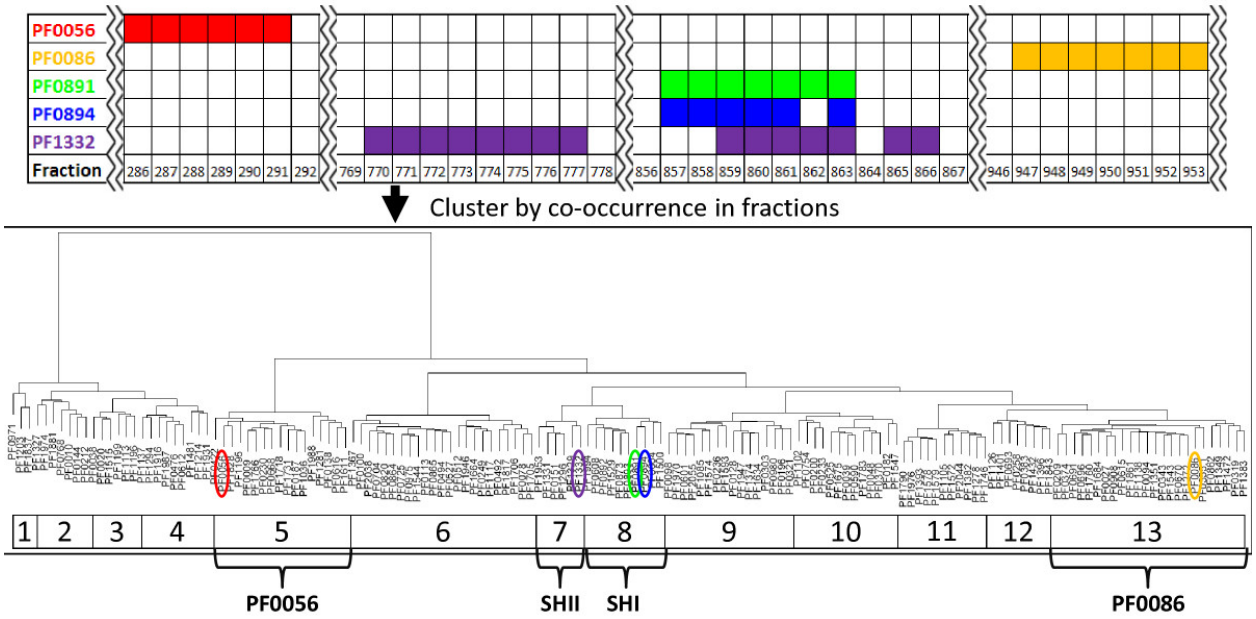
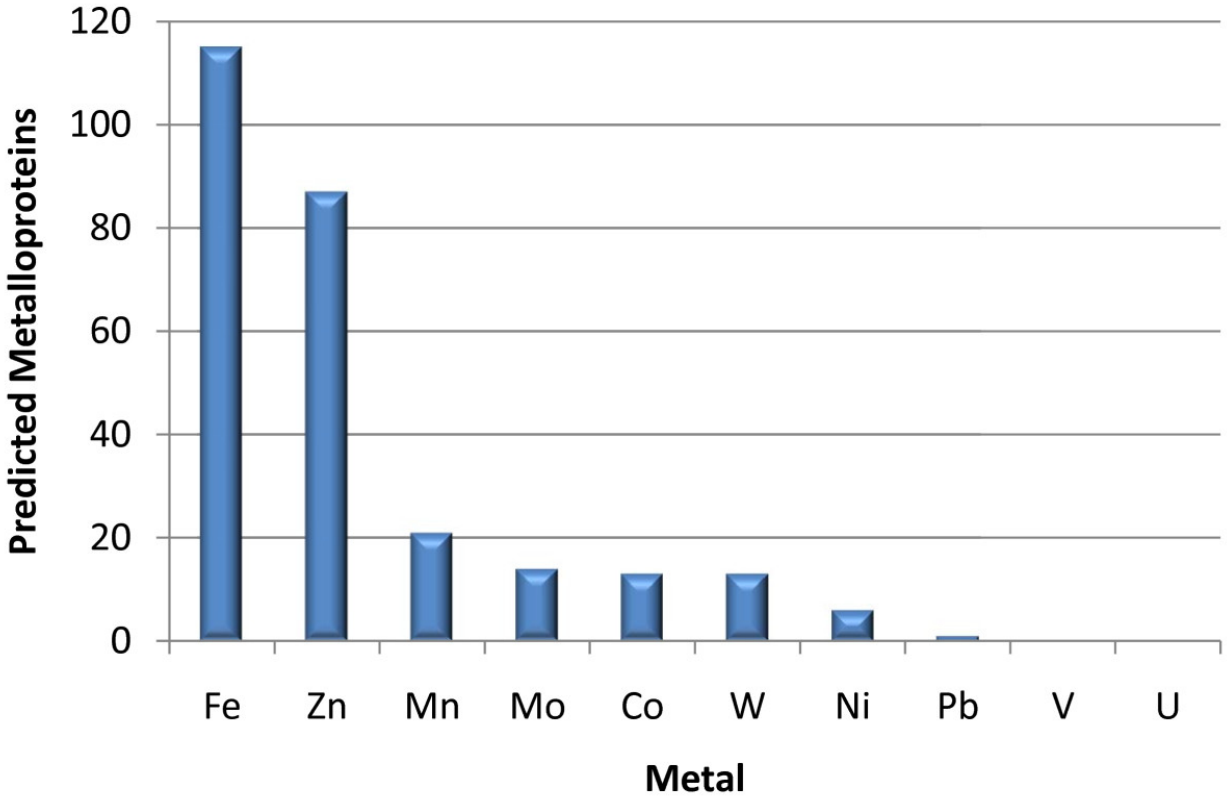

PF0056
PF0086
PF0891
PF0894
PF1332
Fraction
286 287 288 289 290 291 292 769 770 771 772 773 774 775 776 777 778 856 857 858 859 860 861 862 863 864 865 866 867 946 947 948 949 950 951 952 953
Cluster by co-occurrence in fractions
1 2 3 4 5 6 7 8 9 10 11 12 13
PF0056
SHII
SHI
PF0086

Figure 2.5. IPM predicted metalloproteins identified by MS/MS. Out of the 870 proteins identified with two or more peptides, 221 were predicted to be metalloproteins. The majority were predicted to contain Fe or Zn with fewer predicted to contain Mn, Mo, Co, W and Ni.


120
100
80
60
40
20
0
Predicted Metalloproteins
Fe
Zn
Mn
Mo
Co
W
Ni
Pb
V
U
Metal

Additional Files available at:

http://www.biomedcentral.com/1471-2105/12/64/additional

Additional File 1 – Data infrastructure additional details

A more detailed description of the data infrastructure.

(Word doc, 32KB)

Additional File 2 – Relational database schema figure

A diagram illustrating the basic layout of the relational database.

(Powerpoint, ppt, 260KB)

Additional File 3 – OLAP cube figure

A diagram illustrating the basic connections between measure groups and dimensions in the OLAP cube.

(png, 54KB)

Additional File 4 – Relational database schema description

(sql, 500KB)

Additional File 5 – OLAP schema

The complete schema for the OLAP cube as an XML based backup.

(xmla, 1,272KB)

Additional File 6 – Significance curves, function calls for clustering

Significance curves and parameters used to generate all clusterings, clusters. (Powerpoint ppt, 65KB)

Additional File 7 – Cluster Diagrams

Tree diagrams of clusters. The colors are used simply to distinguish the defined clusters. (Adobe pdf, 88KB)

Additional File 8 – Cluster Tables

Supplementary tables 1-8 containing complete cluster details. (Excel xls, 33KB)

Additional File 9 – GMPA scores/occurrences tables

GMPA scores for all metals and proteins studied. The number of fraction occurrences in the data set is listed for each protein as well. (Excel xls, 124KB)

CHAPTER 3

TWO MODEL BIOREMEDIATING-MICROORGANISMS, *DESULFOVIBRIO VULGARIS* AND *PELOSINUS FERMENTANS*, DIFFER FUNDAMENTALLY IN THEIR ABILITIES TO ASSIMILATE AND UTILIZE METALS[2]

[2] Lancaster WA, Menon AL, Scott I, Poole FL, Vaccaro BJ, Geller, J, Hazen TC, Elias DA, Adams MWW. *To be submitted to Metallomics*

ABSTRACT

Although as many as half of all proteins are thought to require a metal cofactor, the metallproteomes of microorganisms remain relatively unexplored. Those from different environments vary greatly in the metals that they assimilate, not just among the metals with well-characterized roles but also those with no known function. Herein we investigated the metal utilization of two microorganisms from very similar natural environments that are of interest because of potential roles in heavy metal detoxification. The metals assimilated and their concentrations in the cytoplasm of *Desulfovibrio vulgaris* strain Hildenborough and *Pelosinus fermentans* strain A11 varied dramatically, with a larger number of metals present in *P. fermentans* and in very different relative amounts. For example, a total of 9 and 19 metals were assimilated into their cytoplasmic fractions, respectively, and *D. vulgaris* did not assimilate zinc or copper whereas *P. fermentans* assimilated both metals. However, bioinformatic analysis of their genome sequences revealed a comparable number of predicted metalloproteins, 899 in *P. fermentans* and 655 in *D. vulgaris*. These allowed some rationalization of the types of metal assimilated in some cases (Fe, Mo, W, V) but not in others (Zn, Cu, Nd, Ce, Pr, Dy, Hf and Th) due to numerous as yet uncharacterized metalloproteins. In addition, both organisms had previously been shown to reduce both chromium (VI) and uranium (VI). We demonstrate here that *P. fermentans*, but not *D. vulgaris* concentrates uranium, but not chromium, into its cytoplasmic fraction. The uranium appears to be present within soluble proteins that await characterization.

INTRODUCTION

Metal ions are essential for all forms of life, and efforts to comprehensively characterize the entire metalloproteome of organisms is a dynamic and emerging field of research (38). It is estimated that one third to one half of all proteins require one or more metal ions to function properly, yet relatively few have been extensively characterized. This is because the amino acid residues in a given protein bind a particular metal site are extremely diverse and similar proteins may bind different metals in different organisms (1). In addition, the study of recombinantly expressed proteins can be misleading since it often leads to the incorporation of a different metal than is found in the natively-produced protein e.g. (10). The need to study natively-produced proteins to reliably determine the true biologically-relevant metal presents another difficulty in that many proteins are expressed at a low level and most organisms do not have genetic systems allowing homologous over-production of putative novel metalloproteins for further characterization (8).

A second aspect of metallomics is the toxicity of certain metals to various microorganisms since relatively little is known about metal toxicity at the molecular level (34, 42). The study of organisms that can tolerate high concentrations of various metals is important not only to our basic understanding of the mechanisms of selective incorporation and detoxification, but also for efforts to harness these abilities for bioremediation. The most commonly encountered metals and metalloids causing adverse health effects in humans are lead, mercury, arsenic and cadmium (18). Uranium and chromium are also acutely toxic elements present at elevated levels in some contaminated natural environments. Efforts are underway to study the microbial communities present at such sites to better understand the diversity and interactions among species in these environments (12, 23, 30). In particular, a great deal of

attention has been focused on the role of sulfate-reducing bacteria of the genus *Desulfovibrio* in the realm of metal bioremediation (including Fe ,Cr, U, Cu, Mn and Ni: (2, 3, 5-7, 11, 13, 17, 20, 24, 26-28, 31-33, 44)). *D. vulgaris* strain Hildenborough and related species have been shown to immobilize metals (Cr,U) through specific reduction by bifunctional enzymes (6, 13, 32) and through the non-specific reaction of metals with sulfide, the product of dissimilatory sulfate reduction. One of the goals of this project was to assess the metalloproteome of the model bioremediator, *D. vulgaris* strain Hildenborough, and determine the metals that it assimilates.

In addition to *Desulfovibrio* sp., members of the genus *Pelosinus* have been very recently shown to be important metal-reducing members of the microbial community in metal-contaminated environments (12, 30). *P. fermentans*, the first representative of the genus, was described as a gram negative, iron–reducing, spore-forming heterotroph (37). It grows on a variety of simple and complex carbon substrates. *Pelosinus* was also found to be the dominant species in lactate-iron enrichment cultures of uranium- and chromium-contaminated soils from various depths suggesting a high degree of environmental and metabolic flexibility (14). A second species, *P. defluvii*, was discovered in groundwater contaminated with halogenated solvents (29) and another *Pelosinus* strain, UFO1, was isolated from uranium-contaminated sediment. This strain was shown to carry out both redox-dependent and phosphate-dependent reactions that resulted in the formation of intracellular uranium granules as well as extracellular uranium precipitates (35, 36). Interestingly, *Pelosinus*-like species became predominant and were considered to be “key drivers of subsurface geochemical conditions” (12) after injection of emulsified vegetable oil into a uranium-contaminated shallow aquifer that was monitored for nine months to determine the microbial diversity and geochemical changes (12). Similarly, a 95-day environmental enrichment culture of uranium- and chromium-contaminated groundwater

also revealed a predominance of *Pelosinus*-like species (30). In this case a total of 16 isolates were obtained that had greater than 99% similarity at the 16S rRNA level to the type strain *Pelosinus fermentans* R7. Although all 16 were able to reduce Fe(III) and some could reduce uranium (VI) or chromium (IV), only one isolate, A11, was also able to reduce all three. A second goal of the current project was therefore to assess the metalloproteome of the uranium- and chromium-reducing *P. fermentans* A11 and determine the metals that it assimilates in laboratory culture. This assessment was aided by a draft genome sequence of *P. fermentans* strain A11 (AKVM00000000) with an overall size of 5 Mb containing 134 contigs and 4758 protein-encoding genes.

An experimental methodology to determine the metals that are assimilated by microorganisms was recently developed and applied to three microorganisms from very different environments (8). These were *Escherichia coli*, a mesophilic bacterium, and two members of the Archaea, *Pyrococcus furiosus*, a neutrophilic hyperthermophile, and *Sulfolobus solfataricus*, an acidophilic thermophile. *E. coli* is found in the gut of humans and other animals and is a facultative anaerobe, *P. furiosus* is found in shallow marine hydrothermal vents and is an obligate anaerobe, and S. *solfataricus* was isolated from volcanic hot springs and is an obligate aerobe. The results of this study showed the assimilation of many unexpected metals by all three organisms and revealed a great diversity in the types and amounts of metals incorporated. For example, under their standard growth conditions, *P. furiosus* was found to assimilate vanadium, lead and uranium, *E. coli* assimilated vanadium, arsenic, lead, cadmium and uranium, while *S. solfataricus* contained barium, lead, tin, antimony and thallium. All of these metals were present as contaminants in the normal growth media of the organism. In the present study we sought to compare the metal utilization and assimilation of two microorganisms that inhabit similar soil

environments, the bioremediators, *D. vulgaris* strain Hildenborough (DvH) and *P. fermentans* strain A11 (Pf A11). Studying the metal assimilation of these two enivonmentally-similar but phylogenetically distinct model organisms will give a new perspective on species-specific metal metabolism and bioremediation, and guide future research into identifying the specific pathways and enzymes involved in metal reduction and sequestration.

MATERIALS & METHODS

Growth.

Both organisms were grown under strictly anaerobic conditions using argon as the headspace. DvH was grown in a modified LS4D lactate sulfate medium as described (7) except that the trace metal (6.25 ml/l instead of 12.5 ml/l) and PIPES (5 mM instead of 30 mM) contents were reduced. DvH was grown in a 600-liter metal fermentor in duplicate 5-liter glass fermentors with no wetted metal parts. Pf A11 was grown in a 600-liter fermentor in the CCM medium previously described (7) except sulfate was omitted and 30 mM fructose replaced lactate as the carbon source. Growth of Pf A11 in Cr and U supplemented media were performed in 4-liter sealed glass bottles using the same fructose CCM medium. In all cases cells were harvested in the mid-log phase (a Sharples centrifuge was used for the 600-liter scale) and the resulting cell mass was flash frozen with liquid nitrogen and stored at -80ºC.

Sampling for metallomics analyses.

Duplicate aliquots (50 ml) of cultures were transferred into 50ml falcon tubes at ~2.5 h time intervals post-innoculation for DvH and ~2 hours for Pf A11, together with samples of the pre-inoculation media. One ml aliquots of all samples were used to measure the optical density

(at 600 nm) while the remaining sample was centrifuged for 15 min at 2800 x g. A sample (15 ml) of the cell-free spent medium supernatant was transferred into acid-washed 15 mL falcon tubes and stored at -20ºC. The cell pellet from one of the duplicate 50 ml tubes was drained well and stored at -20ºC for protein estimation. The remaining pellet was re-suspended in 1-2 ml of the base salt solution for the corresponding growth medium described above, transferred to microfuge tubes and re-centrifuged at 4000 rpm for 4 min to sediment the washed cells. The washed cells and supernatant were stored at -20ºC. The spent medium samples were analyzed for metal content via inductively coupled plasma mass spectrometry (ICP-MS) as described below.

Preparation of washed-cytoplasmic extracts.

All steps were carried out under anaerobic conditions and all glassware was acid-washed. Approximately 10 g (wet weight) of DvH and PfA11 frozen cells were transferred into an anaerobic chamber and 30 ml of anaerobic cell lysis/wash buffer (50 mM Tris HCl, pH 8.0) was added to each. The cell suspensions were stirred until completely homogeneous, transferred into 120 ml serum bottles, sealed with a rubber stoppers, removed from the chamber and placed on ice. Each sample was loaded into a pre-cooled (4°C) 35 ml French Press cell, that was pre-washed four times in anaerobic lysis/wash buffer, and lysed under a flow of argon using the high setting and a pressure of 14000 - 16000 psi (800-1000 setting). The cell extracts were collected in 120 ml serum bottles on ice under a flow of argon and the complete process was repeated. The French Press cell and tubing were washed four-times with anaerobic buffer to prevent carry over between different sample types. The resulting whole cell lysates were centrifuged at 140,000g for 75 min at 4°C. The supernatant, representing the soluble cytoplasmic fraction

(S100) was collected and a 2 ml sample was removed and placed into a buffer-washed Amicon Ultra-4 centrifugal filter device (3 kDa cutoff), which was centrifuged at 4,300 x g until the volume was reduced to 0.5 ml. The flow-through (FT) was retained for metal analysis. Anaerobic lysis/wash buffer (1.5 ml) was added to the retained S100 sample and this was re-centrifuged and the flow-through of the once-washed S100 (W1) was collected for metal analysis. This procedure was repeated and second (W2) and third (W3) washes were collected for metal analysis along with the final retained S100 sample. This was made up to 2 ml with buffer yielding a 256-fold buffer-exchanged, washed S100 sample (S100W).

Metal analysis.

Metal analysis was carried out as previously described (8). Briefly, samples were diluted in a 2% v/v solution of high-purity trace metal grade HNO_3 in glass distilled deionized water and incubated at 37ºC for 90 min to denature protein and release bound metals. Samples were centrifuged (2,400 x g) for 10 mins before analysis using an Agilent 7500ce octopole ICP-MS with a collision/reaction cell. All samples were prepared in 15ml acid-washed polypropylene tubes. All media and S100 samples were analyzed in triplicate, and chromatographic fractions were analyzed in duplicate. Quantitation was performed via in-line addition of internal standard and the use of an external calibration.

Genome comparison.

The total number of metalloproteins encoded in the genomes of DvH and Pf A11 were predicted as previously described (22). Briefly, the predicted amino acid sequences of all protein-encoding genes annotated in the genomes of DvH (RefSeq NC_002937.3 and

NC_005863.1) and Pf A11 (AKVM01000001-AKVM01000134) were analyzed using EBI's iprscan tool (version 4.8; database release 35, 12/14/2011), a standalone version of InterPro (43). The resulting InterPro (IPR) to protein matches were entered into a relational database for each organism. The abstract for each IPR entry was parsed from the accompanying interpro.xml.gz file, and entered into the corresponding database. The IPR abstracts were searched using regular expression patterns relating to metal ions, metal cofactors and metal binding domains and the results were used to classify which IPR entries may encompass proteins that utilize specific metals. The dataset formed the InterProMetalloprotein Database (IPMD) for each organism and these were utilized to predict likely metalloproteins encoded in each genome.

Cytoplasmic fractionation.

A sample (9.7 ml) of the Pf A11 cytoplasmic fraction (S100) was diluted 1:1 with anaerobic 50 mM Tris/HCL buffer pH 8.0 (buffer A) and loaded onto a pre-packed 5 ml DEAE-Sepharose FF column pre-equilibrated with buffer A using an ÄKTA Prime automated liquid chromatography system (GE Healthcare). Unbound protein was washed from the column with buffer A and bound protein was eluted using a linear gradient from 0 - 500mM NaCl in buffer A. A total of 80 fractions (1 ml) were collected and their metal contents were determined as described above. Protein was estimated using the Bradford method using bovine serum albumin as the standard.

RESULTS AND DISCUSSION

Metal content of the media during growth of DvH and Pf A11

Pf A11 and DvH were grown on a 600-liter scale in a metal fermentor and DvH was also grown in a 5-liter glass fermentor. A total of 9 metals (Fe, W, Mn, Co, Zn, Cu, Mo, Se and Ni) were specifically added to the growth media for DvH and Pf A11 (Mg, Ca, Na, and K were also added). Of the 44 elements analyzed, 18 were found in all three media prior to inoculation (V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Ga, As, Rb, Sr, Mo, Sb, Ba, Hf, W and Th) of which 10 (V, Cr, Sr, Ga, As, Rb, Sb, Ba, Hf, Th), were not specifically added to any medium and arise as contaminants from other media components. Two elements, Cd and Lu, were found in the two DvH but not Pf A11 media, while Nd and Tm were found only in the Pf A11 medium. The medium in the metal fermentor for DvH also contained Rh, Sn, Cs, La, Ce, Pr, Sm, Eu, Gd, Ho, Tl and U in detectable amounts to give a total metal count of 32. Differences were observed between the two types of fermentor used for DvH as the glass fermentor medium uniquely contained Pd, Ta, and Pb, to give a total metal count of 23. The total number of detectable metals in the Pf A11 medium was 20. A full list of metal concentrations for each time point in all four cultures is given in Supplementary Table S1.

To assess the changes in available extracellular metals throughout growth of DvH and Pf A11, samples of the cultures were removed periodically, the cellular material was removed by centrifugation and the metal content of the supernatant was determined. The concentrations of several metals decreased during the growth of both organisms, with the effect being much more pronounced for DvH even though their growth rates and cell yields were comparable. For example, in the 600-liter fermentor, DvH and Pf A11 exhibited doubling times of 3.0 and 1.5 hr^{-1}, respectively, and cell yields (wet weight) of 350 g and 725 g, respectively. The initial

concentrations and percent decrease of several metals in the media at the end of growth is summarized in Table 1. Of these metals, ten were decreased to less than 70% of their initial values by DvH in both glass and metal fermentors (Zn, Fe, Co, Mo, Cu, Sb, As, Pb, Sn, Cd), while only Fe decreased by a comparable amount in the Pf A11 growth. Decreases in metal concentrations over the course of DvH growth were similar for metal and glass fermentors.

The decrease in concentration of metals during the DvH growth is perhaps unsurprising given the production of large amounts of hydrogen sulfide by this sulfate-reducing microorganism since transition metals such as Fe, Cu, Co, Mo and Ni readily form insoluble sulfide complexes. Only the soluble metals in the growth media were analyzed here, and this approach does not distinguish between metal lost by extracellular precipitation or from assimilation in to the cell. Nevertheless, this approach does show that several metals, such as Cd, Pb, Sb, Cu and Ba, show a dramatic reduction in concentration during growth of DvH and are reduced to low or sub-nM concentrations. These might limit growth of the organism towards the end of the growth phase. This might be the case for Cd and Ba since, as discussed below, they are assimilated into DvH cells. Pb, Cu and Sb are not assimilated but it is not clear if this is because DvH does not utilize these metals, or that they are not accessible as insoluble sulfide complexes in the medium. On the other hand, for Pf A11, all metals other than Fe were available at the end of growth at more than 50% of their initial concentrations, and so they are unlikely to be a limiting factor.

Metal content of the cytoplasmic fractions of DvH and Pf A11

The metals detected in the soluble cytoplasmic fractions of DvH and Pf A11 cells are summarized in Table 2. We also determined whether the metals were strongly or weakly bound

in to the cytoplasm to macromolecules of at least 3 kDa in size by washing the cytoplasmic extract with buffer three-times using a nominal 3 kDa filter. Figure 1 shows an example of a metal (Co) that was strongly bound and one that was weakly bound (Rb) and readily removed by multiple washes. The amounts of metal bound in the cytoplasm of the two organisms are given in Figure 2 as a side-by-side comparison, and a full list of metals analyzed are given in Supplementary Table S2. Both DvH and Pf A11 assimilated six metals, Mn, Fe, Co, Ni, Mo and W, into soluble cytoplasmic macromolecular complexes above 3 kDa in size. As shown in Table 2 and Figure 3, more than 95% of the metal assimilated by DvH, regardless of the fermentor vessel used, was Fe. This compares with 62% of the total for Pf A11, which also metabolized high relative amounts of Mn and Zn. For reasons that are not at all clear, even though Zn was specifically added to the growth medium, it was not assimilated by DvH. The same was true for Cu, which was readily assimilated by Pf A11 but not by DvH. In addition to those added to the growth media, several metals present as contaminants were taken up by one or both organisms. The cytoplasm of DvH contained tightly-associated Ga, Cd and Ba, while Pf A11 contained V, Cr, Sr, Ce, Pr, Nd, Dy, Hf, Th, and U. Rb was found in the cytoplasm of both organisms but was weakly bound as it was removed by the cytoplasmic washing procedure.

In addition to the differential uptake of several metals by DvH and Pf A11, the overall metal compositions of their cytoplasms varied greatly. For example, although Fe represented the majority of total cytoplasmic metal for both organisms, only Co and Mn were present in DvH at greater than 0.5% of the total, while Pf A11 contained Zn (20%), Mn (7%), Mo, Co, Ni, and Cu in significant quantities. A comparison of three organisms from very different environments, *P. furiosus*, *E. coli* and *S. solfataricus*, revealed species-specific metal assimilation. Of the three, only *P. furiosus* assimilated W, whereas all three assimilated Pb. In contrast both Pf A11 and

DvH assimilated W and neither assimilated Pb. While Pf A11 assimilated both Cr and U, none of the three previously studied organisms assimilated Cr, while *P. furiosus* and *E. coli* contained U (8). These results demonstrate that distinct differences in species-specific metal assimilation also occur among species inhabiting similar environments, such as DvH and Pf A11.

Metal tolerance and assimilation

While U (0.6 nM) and Cr (81 nM) were detected in the growth medium for DvH in the metal fermentor (but not in the glass fermentor, see Supplementary Table S1), neither element was present in the cytoplasmic extract (Table 1). In contrast, Cr (49 nM) was detected in the growth medium of Pf A11 although U was not, but both metals were present into the cytoplasm in the form of macromolecular complexes (Figures 2 and 3). Since Pf A11 cells can reduce both U(VI) and Cr(VI) when used at concentrations of above 0.18 mM (30), it was of great interest to determine the tolerance of Pf A11 to elevated levels of these environmental contaminants and to determine to what extent they were assimilated into the cytoplasmic fraction. The growth of Pf A11 was affected by concentrations of Cr (chromate) above 400 µM but not by U (uranyl acetate) up to 500 µM (Supplementary Figure S1). Supplementation of the medium with 50 nM U resulted in a dramatic increase of 100-fold in the cytoplasmic concentration, from 0.002 to 0.2 nmol U /mg protein (In contrast, supplementation with 100 nM Cr did not result in a significant increase in the cytoplasmic Cr concentration).

Clearly, Pf A11 assimilates U but does it does not actively assimilate Cr. In fact, the cytoplasmic U concentration continued to increase dramatically with supplementation of the growth medium up to 0.5 µM, and this resulted in a 4,500-fold increase (from 0.002 to 9.0 nmol/mg) in that measured in the absence of medium supplementation (Figure 4). However,

only marginal increases in the amount of intracellular U were seen upon further addition of U to the medium above 0.5 μM (Figure 4). These data are the first to indicate that U uptake by Pf A11 is a potentially a physiological process and that U assimilation is not simply occurring passively. Intracellular accumulation of U by a *Pelosinus* species has been previously reported, in which energy dispersive spectroscopy (EDS) revealed that another strain of *Pelosinus, P. fermentans* UFO1, accumulates granules of insoluble U (35). However, in this case the cells were incubated with high concentration of U (100 μM) compared to that used herein (maximum of 0.5 μM U for U uptake) and so the environmental relevance of the U granules is not known.

To ascertain the nature of the U assimalated by Pf A11, the cytoplasmic extract was subjected to anion-exchange chromatography. All procedures including preparing the extract and running the column were carried out under anaerobic conditions. As shown in Figure 5, the U eluted from the column in at least two overlapping peaks within the protein elution profile, strongly suggesting that the U is tightly bound to at least two U-containing, anionic proteins. This discovery represents the first direct evidence for soluble U-containing macromolecules in this or in any organism.

Genome comparison and predicted metalloproteins

The complete genome sequence of DvH has been published (16) and a draft sequence of the Pf A11 genome recently became available. Some properties of the genes annotated in these genomes are summarized in Table 3, over 74% of which have a predicted rather than an unknown function. We recently published a computational framework for predicting metalloproteins on a genome-wide scale using the InterPro annotation database termed the IPMD (22). Using this approach, 4330 of the current 22,361 InterPro entries (IPR) match genes in the

DvH and Pf A11 genomes. A total of 696 of these matched a keyword related to one of 13 metals that were analyzed (Cd, Co, Cu, Cr, Fe, Pb, Mn, Mo, Ni, W, U, V and Zn). Of these 13, only U had no matches to any of the 22,361 IPR entries. Of the 696, 424 had matches in both organisms, while 179 were unique to Pf A11 and 93 were unique to DvH. Overall, 19% of genes in both genomes were covered by least one IPMD match, with a total of 655 in DvH and 899 in Pf A11 (Table 4). On average, there are 2.5 genes per IPR and 2.9 IPRs per gene for DvH. Likewise, PfA11 has 3 genes per IPR and 2.9 IPR per gene indicating a balanced distribution in the analysis between IPRs and genes as well as between organisms. As expected, Fe and Zn matched the most InterPro entries with 315 and 252 hits respectively.

Iron and Zinc: Although Fe was overwhelmingly the predominant metal in the cytoplasmic fraction of DvH, the two organisms had a similar number of Fe IPMD matches, with 239 in DvH and 262 in Pf A11, corresponding to 360 genes in DvH and 476 in PfA11. DvH has robust mechanisms for Fe uptake including ferrous iron transport (Feo) and ferric uptake regulator (Fur) systems and these are also present in the Pf A11 genome. Transcriptomic studies of DvH have shown that Fe-uptake related genes are upregulated under a variety of stresses (15), as cultures transition into stationary phase and nutrients are depleted (7) and in high sulfide conditions (4). The uptake and sequestration of large amounts of Fe by DvH is likely an adaptation to the limited availability of soluble Fe in sulfide-rich natural environments.

Although as yet unstudied, Fe obviously plays a dominant role in the metabolism of Pf A11. In fact, of the 262 Fe IPMD entries for Pf A11, 186 of them are also present in the DvH genome, showing the close metabolic relationship between the two species. From a bioinformatic perspective, a similar situation exists for Zn, where there are 233 matches for Pf

A11 and 175 matches for DvH, with 156 in common. Such data are inconsistent with the fact that Zn, which accounts for 24% of the total metal content of the cytoplasm of Pf A11, was present at an extremely low level in the cytoplasm of DvH. It is possible that some proteins contain Zn in Pf A11 while their homologs in DvH contain Fe, but it seems unlikely that this is prevalent. The low concentration of Zn in the cytoplasm of DvH remains a puzzle.

Molybdenum and Tungsten: Although virtually all characterized organisms metabolize Mo, those that utilize the chemically-analogous element W are much less widespread (19). Moreover, the InterPro entry IPR006655 (molybtopterin oxidoreductase) cannot be used to distinguish between Mo- and W-containing enzymes. Nevertheless, DvH has been shown experimentally to utilize W as both W and Mo are incorporated into the active site of a formate dehydrogenase with a preference for W (9). Hence comparable amounts of W and Mo are found in the cytoplasm of DvH (Figure 2). Pf A11 also contains several genes that could potentially utilize W or Mo and while the relative amount of W is lower than in DvH, its cytoplasmic content (Figure 2) indicates that like DvH, Pf A11 also contains tungstoenzymes.

Manganese, Nickel, Copper and Vanadium: DvH and Pf A11 contain similar amounts of Mn and Ni in their cytoplasmic fractions (Figure 2) and contain comparable numbers of genes that have IPMD matches (Table 4). In particular, both organisms contain Ni-containing hydrogenases, with those from DvH being extensively characterized (25). Both organisms also contain copper-containing oxidoreductasease systems, which in DvH includes oxygen-defensive mechanisms that are even expressed under anaerobic growth conditions (21). In addition, Pf A11, but not DvH, contains homlogs of enzymes of the denitrification pathway, one of which

contains Cu (39). Hence, once more for the two organisms there is a discrepancy between the number of their genes that show an IPMD match with Cu (Table 4), and the relative amounts of copper found in their cytoplasmic fractions (Figure 2). DvH is predicted to harbor many more copper-containing proteins than Pf A11 (37 vs. 27), yet the latter contains more than 20-fold higher content of cytoplasmic Cu. In this case the fact that DvH produces sulfide during growth may be a factor, since Cu readily forms insoluble copper sulfide complexes (40) To what extent Cu availability in a sulfide-rich growth medium limits Cu assimilation by DvH is not known. As shown in Table 4 and Figure 2, genome sequence analyses predicts far more V-related proteins in Pf A11 (36 vs. 7), and its cytoplasm contains 20-fold more V than the cytoplasm of DvH. One explanation for V accumulation in Pf A11 is the presence of a protein matching "Vanadium/alternative nitrogenase delta subunit" (IPR004349) that is not found in DvH, indicating the possible presence of an alternative nitrogenase system utilizing V instead of Mo. Although it is unlikely nitrogenase enzymes are highly expressed under the ammonium abundant growth conditions, Pf A11 may have more robust uptake mechanisms to supply this pathway.

Uranium and Chromium: No genes in either DvH or Pf A11 match IPMD keywords relating to U. Only DvH had IPMD matches relating to Cr and these three IPRs (IPR003370, IPR014047, IPR018634) cover, with some overlap, only three genes: one on the chromosome (DVU0426) and two on the megaplasmid (DVUA0093, DVUA0094). The IPR entries suggest all three genes are related to chromate transport. This lacks of hits for metal reductases reflects the limited knowledge about specific metal reduction mechanisms and the types of metalloproteins that might be involved. Proteins that have been implicated, such as cytochrome c_3 and periplasmic hydrogenases (11, 41), which are Fe-containing proteins, have other

metabolic roles and are ubiquitous among a variety of related organisms that do not reduce U or Cr. Until specific chelators, reductases, or variations in redox proteins that specifically impart the ability to reduce hexavalent metal ions are identified, bioinformatic analysis alone is unlikely to allow the prediction of such capability from genome sequences. However, the identification of soluble cytoplasmic U complexes presented herein may provide the means to characterize such proteins that can be used to identify homologous proteins in the genomes of other organisms or in the metagenomes of environmental consortia.

Other minor constituents: As with U and Cr, the other metals that are present in the cytoplasm of Pf A11 but not in DvH, such as Nd, Ce, Pr, Dy, Hf and Th, cannot be rationalized in terms of bioinformatic analysis since there are no known proteins that naturally contain such metals, and no structural or catalytic roles for them are present in the databases such as InterPro. As with U and Cr, native biomass of these organisms must be analyzed to provide insights into what cannot be gained from bioinformatic approaches.

CONCLUSIONS

Metallomic investigation of two species of interest to bioremediation efforts, DvH and Pf A11, revealed fundamental differences in their abilities to assimilate metals. A greater diversity in the types of intracellular metal was seen in Pf A11. For example, it assimilates V, Cu, Zn, Sn and U, but DvH does not. Cytoplasmic metal and bioinformatic analysis suggest that both Pf A11 and DvH use W as well as Mo-containing oxidoreductases and both are predicted to contain comparable numbers of Cu- and Zn-containing enzymes, yet the cytoplasmic concentrations of these metals in DvH was dramatically lower than those found in Pf A11. Both of these organisms

have been investigated for their ability to reduce the toxic metals Cr and U, but neither accumulated Cr in their cytoplasmic fractions while Pf A11, but not DvH, assimilated U. Moreover evidence is presented to show that the U is actively incorporated into macromolecular complexes that could be separated by column chromatography and are assumed to be U-containing proteins. A related *Pelosinus* species, UFO1, has been shown to accumulate insoluble U-containing granules by unknown mechanisms, but U concentrations in that study were 200-times the highest used here (35). This study provides the first evidence for soluble intracellular U species bound to macromolecular complexes in this species. The ability of *Pelosinus* to acquire high intracellular concentrations of soluble U-containing macromolecular species without apparent toxic effects makes it a particularly attractive model organism for understanding transport, sequestration and detoxification mechanisms of this environmental contaminant. Additional studies of these recently discovered *Pelosinus* strains, which have been shown to be highly abundant in heavy metal-contaminated environments (12, 30), will improve our understanding of the role of toxic metals in microbial communities.

ABBREVIATIONS

DvH	*Desulfovibrio vulgaris* Hildenborough
DEAE	Diethylaminoethyl anion exchange chromatography
DUF	Domain of unknown function
FT	Flow-through
ICP-MS	Inductively couples plasma mass spectrometry
Pf A11	*Pelosinus fermentans* strain A11
S100	Supernatant after centrifugation of a cell-free extract at 100,000 x g

S100W S100 after washing three times using a 3 kDa filter

UCP Uncharacterized conserved protein

UPF Uncharacterized domains

ACKNOWLEDGEMENTS

This work conducted by ENIGMA- Ecosystems and Networks Integrated with Genes and Molecular Assemblies (http://enigma.lbl.gov), a Scientific Focus Area Program at Lawrence Berkeley National Laboratory, was supported by the Office of Science, Office of Biological and Environmental Research, of the U. S. Department of Energy under Contract No. DE-AC02-05CH11231.

REFERENCES

1. **Bertini, I., and G. Cavallaro.** 2010. Bioinformatics in bioinorganic chemistry. Metallomics **2:**39-51.

2. **Biswas, K. C., N. A. Woodards, H. Xu, and L. L. Barton.** 2009. Reduction of molybdate by sulfate-reducing bacteria. Biometals **22:**131-9.

3. **Cabrera, G., R. Perez, J. M. Gomez, A. Abalos, and D. Cantero.** 2006. Toxic effects of dissolved heavy metals on *Desulfovibrio vulgaris* and *Desulfovibrio* sp. strains. J Hazard Mater **135:**40-6.

4. **Caffrey, S. M., and G. Voordouw.** 2010. Effect of sulfide on growth physiology and gene expression of *Desulfovibrio vulgaris* Hildenborough. Antonie Van Leeuwenhoek **97:**11-20.

5. **Chardin, B., A. Dolla, F. Chaspoul, M. L. Fardeau, P. Gallice, and M. Bruschi.** 2002. Bioremediation of chromate: thermodynamic analysis of the effects of Cr(VI) on sulfate-reducing bacteria. Appl Microbiol Biotechnol **60:**352-60.

6. **Chardin, B., M. T. Giudici-Orticoni, G. De Luca, B. Guigliarelli, and M. Bruschi.** 2003. Hydrogenases in sulfate-reducing bacteria function as chromium reductase. Appl Microbiol Biotechnol **63:**315-21.

7. **Clark, M. E., Q. He, Z. He, K. H. Huang, E. J. Alm, X. F. Wan, T. C. Hazen, A. P. Arkin, J. D. Wall, J. Z. Zhou, and M. W. Fields.** 2006. Temporal transcriptomic analysis as *Desulfovibrio vulgaris* Hildenborough transitions into stationary phase during electron donor depletion. Appl Environ Microbiol **72:**5578-88.

8. **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.** 2010. Microbial metalloproteomes are largely uncharacterized. Nature **466:**779-82.

9. **da Silva, S. M., C. Pimentel, F. M. Valente, C. Rodrigues-Pousada, and I. A. Pereira.** 2011. Tungsten and molybdenum regulation of formate dehydrogenase expression in *Desulfovibrio vulgaris* Hildenborough. J Bacteriol **193:**2909-16.

10. **Eidsness, M. K., S. E. O'Dell, D. M. Kurtz, Jr., R. L. Robson, and R. A. Scott.** 1992. Expression of a synthetic gene coding for the amino acid sequence of *Clostridium pasteurianum* rubredoxin. Protein Eng **5:**367-71.

11. **Elias, D. A., J. M. Suflita, M. J. McInerney, and L. R. Krumholz.** 2004. Periplasmic cytochrome c3 of *Desulfovibrio vulgaris* is directly involved in H2-mediated metal but not sulfate reduction. Appl Environ Microbiol **70:**413-20.

12. **Gihring, T. M., G. Zhang, C. C. Brandt, S. C. Brooks, J. H. Campbell, S. Carroll, C. S. Criddle, S. J. Green, P. Jardine, J. E. Kostka, K. Lowe, T. L. Mehlhorn, W. Overholt, D. B. Watson, Z. Yang, W. M. Wu, and C. W. Schadt.** 2011. A limited microbial consortium is responsible for extended bioreduction of uranium in a contaminated aquifer. Appl Environ Microbiol **77:**5955-65.

13. **Goulhen, F., A. Gloter, F. Guyot, and M. Bruschi.** 2006. Cr(VI) detoxification by *Desulfovibrio vulgaris* strain Hildenborough: microbe-metal interactions studies. Appl Microbiol Biotechnol **71:**892-7.

14. **Hansel, C. M., S. Fendorf, P. M. Jardine, and C. A. Francis.** 2008. Changes in bacterial and archaeal community structure and functional diversity along a geochemically variable soil profile. Appl Environ Microbiol **74:**1620-33.

15. **He, Z., A. Zhou, E. Baidoo, Q. He, M. P. Joachimiak, P. Benke, R. Phan, A. Mukhopadhyay, C. L. Hemme, K. Huang, E. J. Alm, M. W. Fields, J. Wall, D. Stahl, T. C. Hazen, J. D. Keasling, A. P. Arkin, and J. Zhou.** 2009. Global transcriptional, physiological, and metabolite analyses of the responses of *Desulfovibrio vulgaris* hildenborough to salt adaptation. Appl Environ Microbiol **76:**1574-86.

16. **Heidelberg, J. F., R. Seshadri, S. A. Haveman, C. L. Hemme, I. T. Paulsen, J. F. Kolonay, J. A. Eisen, N. Ward, B. Methe, L. M. Brinkac, S. C. Daugherty, R. T. Deboy, R. J. Dodson, A. S. Durkin, R. Madupu, W. C. Nelson, S. A. Sullivan, D. Fouts, D. H. Haft, J. Selengut, J. D. Peterson, T. M. Davidsen, N. Zafar, L. Zhou, D. Radune, G. Dimitrov, M. Hance, K. Tran, H. Khouri, J. Gill, T. R. Utterback, T. V. Feldblyum, J. D. Wall, G. Voordouw, and C. M. Fraser.** 2004. The genome sequence of the anaerobic, sulfate-reducing bacterium *Desulfovibrio vulgaris* Hildenborough. Nat Biotechnol **22:**554-9.

17. **Humphries, A. C., K. P. Nott, L. D. Hall, and L. E. Macaskie.** 2004. Continuous removal of Cr(VI) from aqueous solution catalysed by palladised biomass of *Desulfovibrio vulgaris*. Biotechnol Lett **26:**1529-32.

18. **Jarup, L.** 2003. Hazards of heavy metal contamination. Br Med Bull **68:**167-82.

19. **Kletzin, A., and M. W. Adams.** 1996. Tungsten in biological systems. FEMS Microbiol Rev **18:**5-63.

20. **Klonowska, A., M. E. Clark, S. B. Thieman, B. J. Giles, J. D. Wall, and M. W. Fields.** 2008. Hexavalent chromium reduction in *Desulfovibrio vulgaris* Hildenborough causes transitory inhibition of sulfate reduction and cell growth. Appl Microbiol Biotechnol **78:**1007-16.

21. **Lamrabet, O., L. Pieulle, C. Aubert, F. Mouhamar, P. Stocker, A. Dolla, and G. Brasseur.** 2011. Oxygen reduction in the strict anaerobe *Desulfovibrio vulgaris* Hildenborough: characterization of two membrane-bound oxygen reductases. Microbiology **157:**2720-32.

22. **Lancaster, W. A., J. L. Praissman, F. L. Poole, 2nd, A. Cvetkovic, A. L. Menon, J. W. Scott, F. E. Jenney, Jr., M. P. Thorgersen, E. Kalisiak, J. V. Apon, S. A. Trauger, G. Siuzdak, J. A. Tainer, and M. W. Adams.** 2011. A computational framework for proteome-wide pursuit and prediction of metalloproteins using ICP-MS and MS/MS data. BMC Bioinformatics **12:**64.

23. **Lin, X., J. McKinley, C. T. Resch, R. Kaluzny, C. L. Lauber, J. Fredrickson, R. Knight, and A. Konopka.** 2012. Spatial and temporal dynamics of the microbial community in the Hanford unconfined aquifer. ISME J.

24. **Lovley, D. R., P. K. Widman, J. C. Woodward, and E. J. Phillips.** 1993. Reduction of uranium by cytochrome c3 of *Desulfovibrio vulgaris*. Appl Environ Microbiol **59:**3572-6.

25. **Marques, M. C., R. Coelho, A. L. De Lacey, I. A. Pereira, and P. M. Matias.** 2009. The three-dimensional structure of [NiFeSe] hydrogenase from *Desulfovibrio vulgaris* Hildenborough: a hydrogenase without a bridging ligand in the active site in its oxidised, "as-isolated" state. J Mol Biol **396:**893-907.

26. **Martins, M., M. L. Faleiro, R. J. Barros, A. R. Verissimo, M. A. Barreiros, and M. C. Costa.** 2009. Characterization and activity studies of highly heavy metal resistant sulphate-reducing bacteria to be used in acid mine drainage decontamination. J Hazard Mater **166:**706-13.

27. **Martins, M., M. L. Faleiro, S. Chaves, R. Tenreiro, and M. C. Costa.** 2010. Effect of uranium (VI) on two sulphate-reducing bacteria cultures from a uranium mine site. Sci Total Environ **408:**2621-8.

28. **Martins, M., M. L. Faleiro, S. Chaves, R. Tenreiro, E. Santos, and M. C. Costa.** 2010. Anaerobic bio-removal of uranium (VI) and chromium (VI): comparison of microbial community structure. J Hazard Mater **176:**1065-72.

29. **Moe, W. M., R. E. Stebbing, J. U. Rao, K. S. Bowman, M. F. Nobre, M. S. da Costa, and F. A. Rainey.** 2011. *Pelosinus defluvii* sp. nov., isolated from chlorinated solvent contaminated groundwater, emended description of the genus *Pelosinus*, and transfer of *Sporotalea propionica* to *Pelosinus propionicus* comb. nov. Int J Syst Evol Microbiol.

30. **Mosher, J. J., T. J. Phelps, M. Podar, R. A. Hurt, Jr., J. H. Campbell, M. M. Drake, J. G. Moberly, C. W. Schadt, S. D. Brown, T. C. Hazen, A. P. Arkin, A. V. Palumbo, B. A. Faybishenko, and D. A. Elias.** 2012. Microbial community succession during lactate amendment and electron acceptor limitation reveals a predominance of metal-reducing *Pelosinus* spp. Appl Environ Microbiol **78:**2082-91.

31. **Park, H. S., S. Lin, and G. Voordouw.** 2008. Ferric iron reduction by *Desulfovibrio vulgaris* Hildenborough wild type and energy metabolism mutants. Antonie Van Leeuwenhoek **93:**79-85.

32. **Payne, R. B., D. M. Gentry, B. J. Rapp-Giles, L. Casalot, and J. D. Wall.** 2002. Uranium reduction by *Desulfovibrio desulfuricans* strain G20 and a cytochrome c3 mutant. Appl Environ Microbiol **68:**3129-32.

33. **Pietzsch, K., and W. Babel.** 2003. A sulfate-reducing bacterium that can detoxify U(VI) and obtain energy via nitrate reduction. J Basic Microbiol **43:**348-61.

34. **Piotrowska-Seget, Z., M. Cycoń, and J. Kozdrój.** 2005. Metal-tolerant bacteria occurring in heavily polluted soil and mine spoil. Applied Soil Ecology **28:**237-246.

35. **Ray, A. E., J. R. Bargar, V. Sivaswamy, A. C. Dohnalkova, Y. Fujita, B. M. Peyton, and T. S. Magnuson.** 2011. Evidence for multiple modes of uranium immobilization by an anaerobic bacterium. Geochimica et Cosmochimica Acta **75:**2684-2695.

36. **Ray, A. E., S. A. Connon, P. P. Sheridan, J. Gilbreath, M. Shields, D. T. Newby, Y. Fujita, and T. S. Magnuson.** 2010. Intragenomic heterogeneity of the 16S rRNA gene in strain UFO1 caused by a 100-bp insertion in helix 6. FEMS Microbiol Ecol **72:**343-53.

37. **Shelobolina, E. S., K. P. Nevin, J. D. Blakeney-Hayward, C. V. Johnsen, T. W. Plaia, P. Krader, T. Woodard, D. E. Holmes, C. G. Vanpraagh, and D. R. Lovley.** 2007. *Geobacter pickeringii* sp. nov., *Geobacter argillaceus* sp. nov. and *Pelosinus fermentans* gen. nov., sp. nov., isolated from subsurface kaolin lenses. Int J Syst Evol Microbiol **57:**126-35.

38. **Szpunar, J.** 2005. Advances in analytical methodology for bioinorganic speciation analysis: metallomics, metalloproteomics and heteroatom-tagged proteomics and metabolomics. Analyst **130:**442-65.

39. **Tavares, P., A. S. Pereira, J. J. Moura, and I. Moura.** 2006. Metalloenzymes of the denitrification pathway. J Inorg Biochem **100:**2087-100.

40. **Thompson, R. A., and G. R. Helz.** 1994. Copper speciation in sulfidic solutions at low sulfur activity: further evidence for cluster complexes? Geochim. Cosmochim. Acta **58:**2971-83.

41. **Wall, J. D., and L. R. Krumholz.** 2006. Uranium reduction. Annu Rev Microbiol **60:**149-66.

42. **Wysocki, R., and M. J. Tamas.** 2010. How Saccharomyces cerevisiae copes with toxic metals and metalloids. FEMS Microbiol Rev **34:**925-51.

43. **Zdobnov, E. M., and R. Apweiler.** 2001. InterProScan--an integration platform for the signature-recognition methods in InterPro. Bioinformatics **17:**847-8.

44. **Zhang, W., D. E. Culley, L. Nie, and J. C. Scholten.** 2007. Comparative transcriptome analysis of *Desulfovibrio vulgaris* grown in planktonic culture and mature biofilm on a steel surface. Appl Microbiol Biotechnol **76:**447-57.

Table 3.1. Metal content of the growth media for DvH and Pf A11 and the decrease in concentration at the end of cell growth. DvH was grown in a metal and in a glass fermentor as indicated. The percent decrease is the concentration of each metal in the medium at the end of growth (mid-log phase) compared to the initial concentration prior to inoculation.

	Pf A11		Glass DvH		Metal DvH	
	Concentration (nM)	Decrease (%)	Concentration (nM)	Decrease (%)	Concentration (nM)	Decrease (%)
Zn	2886.9	47.3	13712.2	87.9	8257.4	97.4
Fe	905.6	83.9	38548.0	91.1	17079.2	87.8
Mn	736.8	34.5	17712.6	8.2	14903.9	21.2
Co	521.0	33.3	8555.7	89.3	7138.3	98.3
Mo	302.6	39.4	1405.3	90.3	1110.0	93.5
Ni	187.2	32.5	2592.7	33.2	2141.5	61.9
W	165.9	42.2	186.4	21.9	244.0	18.2
Sr	162.5	33.8	256.7	9.9	116.5	10.6
Ba	143.4	38.4	77.4	21.5	103.1	64.2
Rb	140.8	34.6	293.5	6.4	98.6	24.5
Cu	104.9	56.0	806.9	97.9	78.7	100.0
Cr	49.1	39.2	51.4	14.8	81.0	18.6
V	40.1	31.3	9.5	0.3	16.5	33.5
Sb	3.6	40.8	5.4	100.0	3.4	71.1
As	2.5	51.1	8.7	23.1	6.9	72.3
Pb	ND	-	29.4	97.1	1.1	88.7
Sn	ND	-	ND	-	5.9	99.0
Cd	ND	-	48.9	98.4	2.5	76.0
Cs	ND	-	ND	-	8.0	26.0
U	ND	-	ND	-	0.6	28.1

Table 3.2. Metals strongly and weakly (-) bound to macromolecules (> 3 kDa) in the cytoplasm of DvH and Pf A11. The concentrations of the metals shown are before the samples were washed three times with buffer as described in the Materials and Methods. Metals whose concentration decreased by more than 50% because of the washing procedure were designated weakly bound and are so indicated (-).

	Pf A11	DvH
Fe	3844.27	29717.59
Zn	1454.63	ND
Mn	433.27	640.61
Mo	169.54	71.04
Ni	109.41	156.78
Co	69.81	660.82
Cu	51.69	ND
Ba	18.84	6.60
W	11.60	133.81
Cr	9.36	ND
Ga	7.45	2.59
V	4.46	ND
Sr	2.65	4.21(-)
Nd	2.15	ND
Rb	0.73(-)	5.82(-)
Sn	0.52	ND
Ce	0.30	ND
Dy	0.27	ND
Cd	0.12	0.90
Th	0.09	ND
Hf	0.09	ND
U	0.05	ND
Pr	0.02	ND

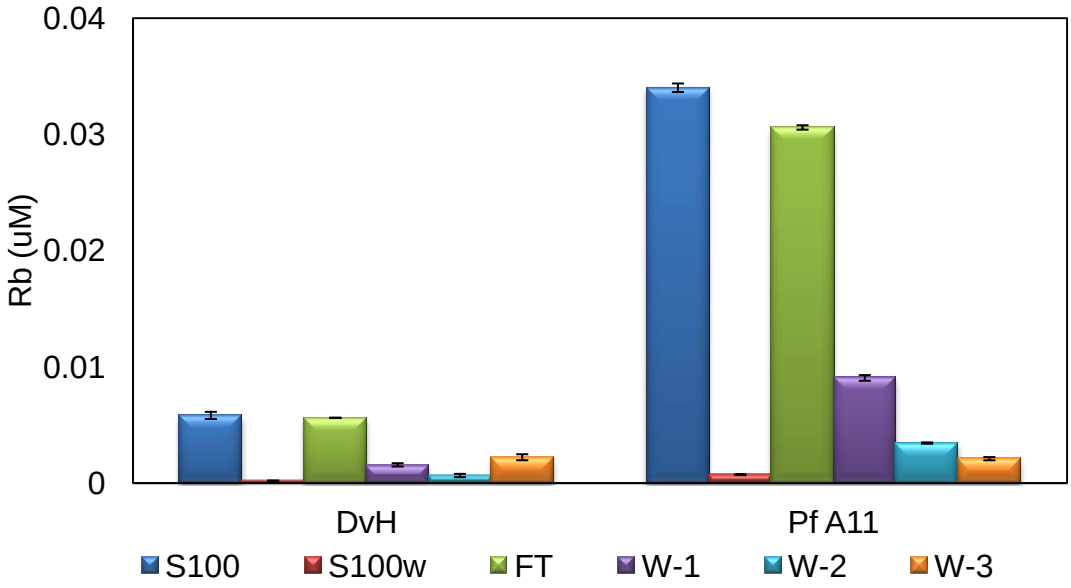
Table 3.3. Comparison of the genome sequences of DvH (complete) and Pf A11 (draft). The DvH chromosome is 3.6 Mb (RefSeq NC_002937.3) with a 200Kb megaplasmid (RefSeq NC_005863.1). The draft genome sequence of *P. fermentans* strain A11 is 5.2 Mb in size (AKVM01000001-AKVM01000134). The number of predicted protein-encoding genes is given, along with the number that match at least one InterPro entry and the number of those that match "hypothetical protein" IPRs or Domain of unknown function (DUF), Uncharacterized conserved protein (UCP), or Uncharacterized domain (UPF).

	Contigs	Nucleotides (Mb)	Genes	Genes with IPR match	Hypothetical proteins (IPR)	IPR DUF, UCP, or UPF
DvH	2	3.8	3536	2605 (74%)	129	55
Pf A11	124	5.2	4758	3921 (82%)	210	62

Table 3.4. Genome-wide comparison of the predicted metalloproteomes of DvH and Pf A11. The number of genes with an IPMD match is given for each organism. Some genes have matches with multiple metals, and a total of 655 DvH and 899 Pf A11 genes had a match to at least one metal.

	DvH	Pf A11
Cadmium	20	24
Cobalt	106	153
Copper	37	27
Chromium	3	0
Iron	360	476
Lead	6	9
Manganese	15	38
Molybdenum	74	114
Nickel	58	72
Tungsten	14	17
Uranium	0	0
Vanadium	7	36
Zinc	202	292

Figure 3.1. Strongly and weakly bound metal in the cytoplasm of Pf A11 and DvH. Cytoplasmic extracts were prepared (S100) and each was washed three-times using a 3 kDa filter as described in the Materials and Methods. The Co and Rb concentrations of the washed extract (S100w), the initial flow through (FT), and the flow after the three washes (W-1, -2 and -3).


0.04
0.03
0.02
0.01
0
Rb (uM)
DvH
Pf A11
S100
S100w
FT
W-1
W-2
W-3

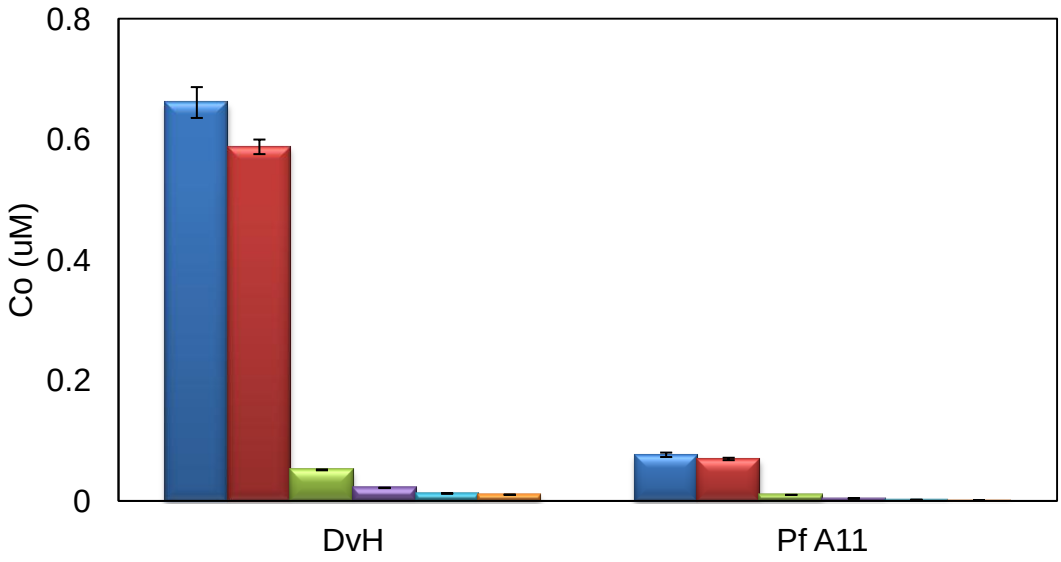

0.8
0.6
0.4
0.2
0
Co (uM)
DvH
Pf A11

Figure 3.2. Metals that are strongly-bound to macromolecules (> 3 kDa) in cytoplasmic extracts (S100) of DvH and/or Pf A11. The cytoplasmic extracts were prepared and each was washed three-times as described in the Materials and Methods prior to metal analysis.

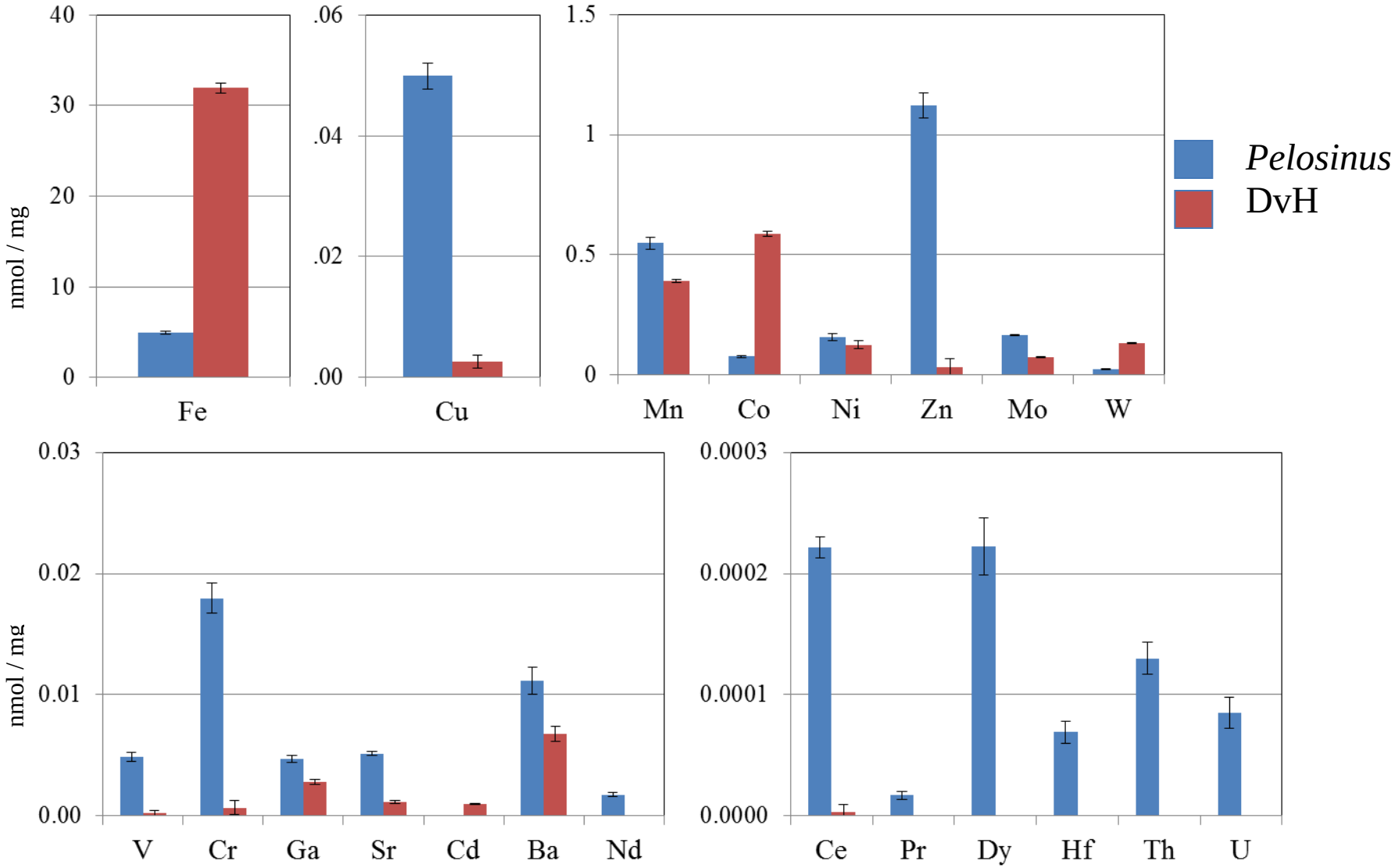
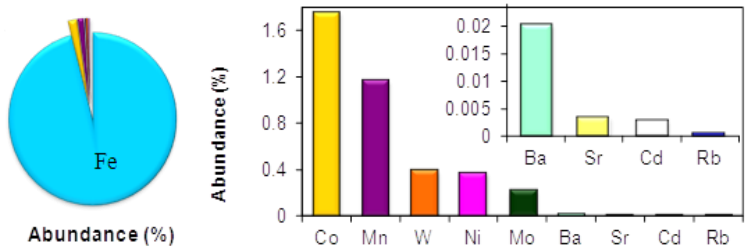
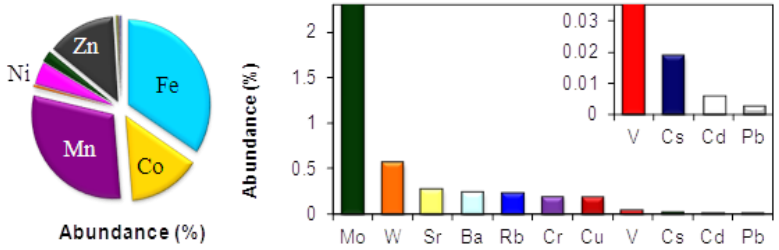

nmol / mg
40
30
20
10
0
Fe
.06
.04
.02
.00
Cu
1.5
1
0.5
0
Mn
Co
Ni
Zn
Mo
W
Pelosinus
DvH
nmol / mg
0.03
0.02
0.01
0.00
V
Cr
Ga
Sr
Cd
Ba
Nd
0.0003
0.0002
0.0001
0.0000
Ce
Pr
Dy
Hf
Th
U

Figure 3.3. Comparison of the metal composition of the growth media and the cytoplasmic fractions (S100W) of DvH and Pf A11 grown in 600-liter cultures.

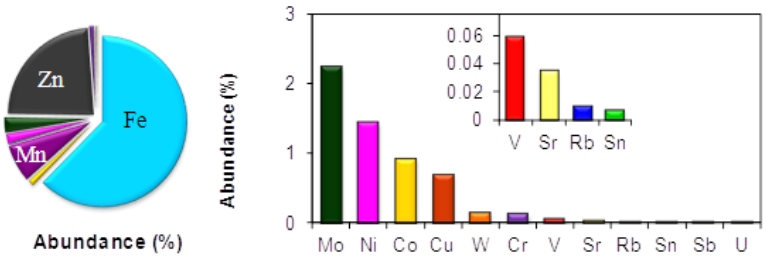
DvH Washed Cytoplasm



Growth Medium (Defined Lactate-Sulfate)



Pelosinus A11 Washed Cytoplasm



Growth Medium (Defined Fructose)

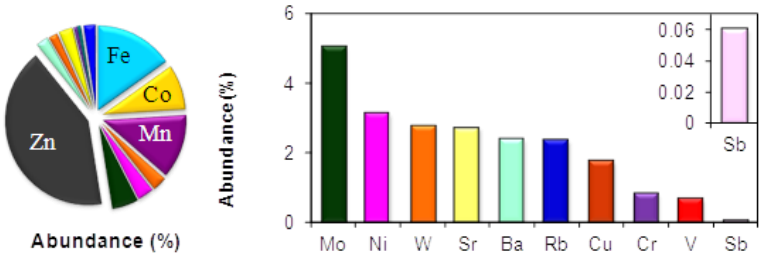


Figure 3.4. Strongly-bound uranium (> 3 kDa macromolecules) in the cytoplasmic fraction (S100W) of Pf A11 cells grown in media supplemented with uranium.

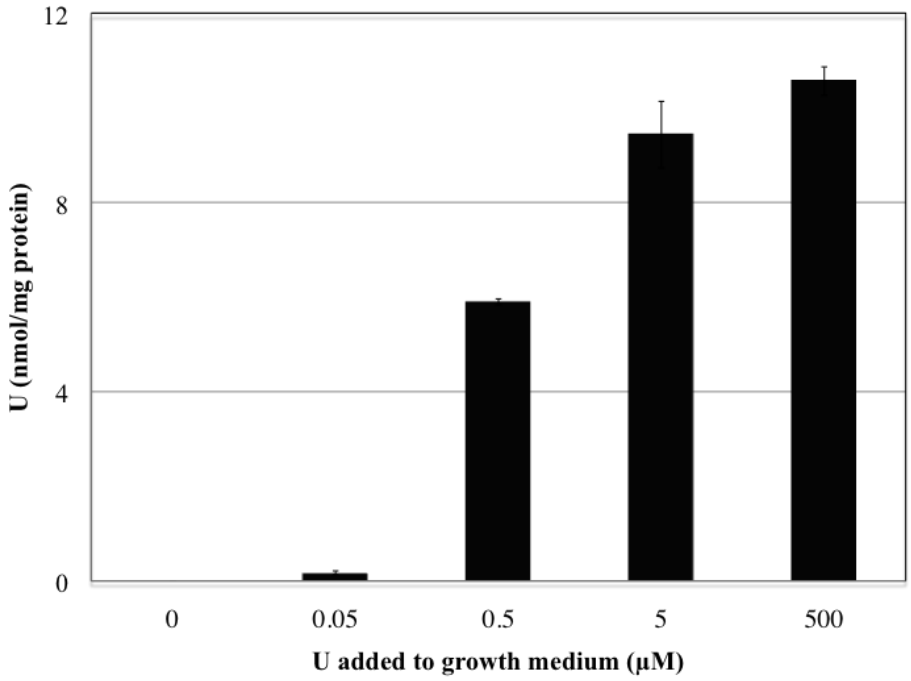
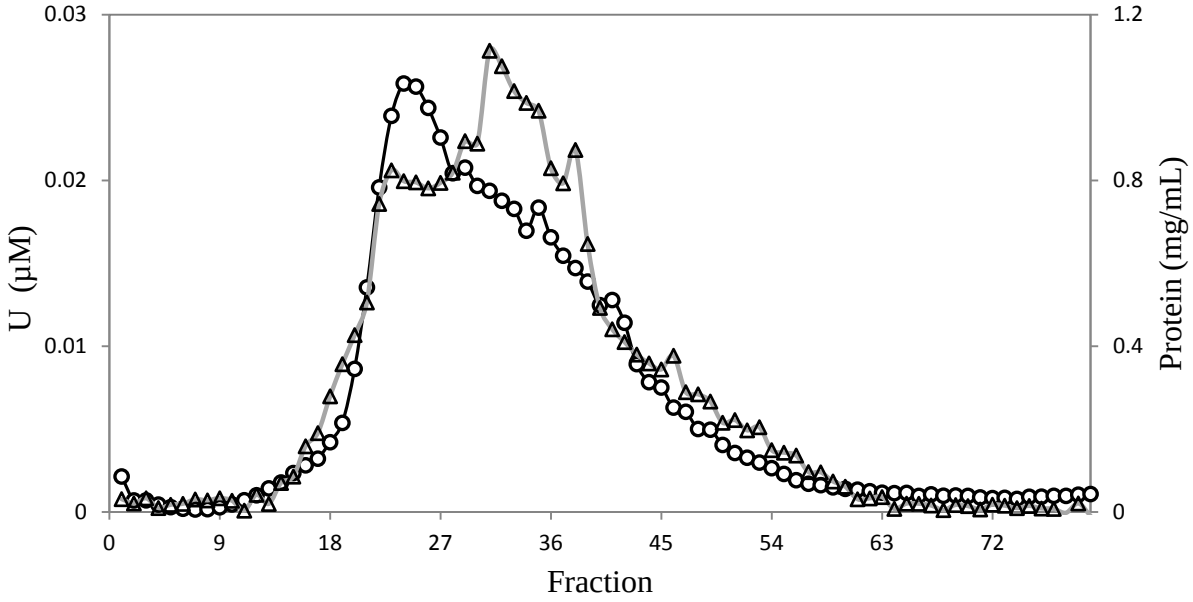

12
8
4
0
U (nmol/mg protein)
0
0.05
0.5
5
500
U added to growth medium (µM)

Figure 3.5. Elution profile of protein and uranium after anion-exchange chromatography of a cytoplasmic extract of Pf A11 from cells grown in a supplemented medium. The concentrations of uranium (circles) and protein (triangles) in each fraction are indicated.


0.03
0.02
0.01
0
U (µM)
1.2
0.8
0.4
0
Protein (mg/mL)
0
9
18
27
36
45
54
63
72
Fraction

SUPPORTING MATERIAL

Table 3.S1. The elements in the spent media of the four fermentations (A-D) for each of the 44 measured are given in nM with standard deviations.

A. Elements in the spent medium of 5 l DvH glass reactor 1

	Glass DvH Reactor 1 t0		Glass DvH Reactor 1 t1		Glass DvH Reactor 1 t2		Glass DvH Reactor 1 t3		Glass DvH Reactor 1 t4		Glass DvH Reactor 1 t5	
	nM	**SD**	**nM**	**SD**	**nM**	**SD**	**nM**	**SD**	**nM**	**SD**	**nM**	**SD**
V	8.50	1.61	9.03	1.28	10.03	0.62	8.52	1.04	8.90	0.69	8.74	0.27
Cr53	47.06	4.51	44.10	13.74	47.92	7.99	51.13	2.91	50.45	6.28	44.92	2.94
Mn	16353.41	1225.55	16371.82	1144.60	16480.23	617.97	16601.59	750.15	16729.77	633.07	16011.14	489.41
Fe	35464.29	2161.30	31908.48	2347.52	25663.39	1167.58	13510.04	645.40	7125.00	271.61	3314.73	219.46
Co	7931.57	466.54	5256.36	345.60	4112.92	311.66	2575.36	157.74	1675.42	303.72	1059.03	121.82
Ni	2447.25	221.87	2335.06	143.47	2338.94	71.23	2240.94	34.90	2087.31	85.18	1853.75	77.67
Cu	691.55	111.38	260.63	23.40	117.16	24.05	34.39	1.78	18.86	2.02	18.66	3.12
Zn66	14863.07	5705.45	6557.95	630.79	5340.91	3282.88	3730.11	1126.57	2510.51	1009.59	1724.03	780.25
Zn68	12906.07	4925.27	5716.54	531.19	4647.24	2829.52	3230.29	978.19	2195.68	840.26	1518.20	671.11
Ga	28.35	5.01	26.33	1.93	23.68	0.64	24.01	2.08	23.26	0.66	22.12	3.39
As	8.61	0.55	7.40	0.90	8.16	0.95	7.58	0.51	7.38	0.85	6.54	0.95
Rb	276.09	17.07	274.94	22.65	287.91	11.42	282.00	6.65	284.47	7.02	263.38	14.22
Sr	235.36	20.27	229.64	16.54	243.41	9.90	242.68	7.96	236.59	11.75	227.39	17.09
Mo95	1294.07	64.69	951.28	56.05	664.46	37.89	300.47	8.78	171.48	30.11	149.69	13.83
Mo98	1167.82	59.75	856.76	51.31	598.32	36.85	270.88	8.00	152.64	27.49	133.03	12.60
Ru	0.33	0.24	0.30	0.24	0.11	0.09	0.14	0.10	0.01	0.01	0	0
Rh	0	0	0	0	0	0	0	0	0	0	0	0
Pd	216.46	60.78	3178.21	597.75	2090.00	293.00	2369.64	585.17	1506.75	351.73	930.64	244.71
Cd111	44.63	2.76	22.09	1.91	11.83	1.73	5.07	1.47	1.80	0.55	0.84	0.17
Sn120	2.49	1.82	0.27	0.30	0.41	0.46	0	0	0	0	0.03	0.06
Sb	4.84	0.17	4.45	0.24	4.09	0.34	1.60	0.07	0.14	0.23	0	0
Te	2.16	0.60	1.63	0.62	1.20	0.28	1.89	0.55	1.10	0.50	1.66	0.22
Cs	1.51	0.38	1.08	0.05	1.55	0.40	1.65	1.04	1.19	0.13	0.99	0.12
Ba	71.50	5.04	67.23	3.06	65.69	5.60	66.88	1.26	65.30	2.72	63.98	1.95
La	0	0	0	0	0	0	0.69	1.38	0	0	0	0
Ce	0.31	0.05	0.27	0.18	0.52	0.61	1.59	3.17	0.02	0.03	0.07	0.15
Pr	0	0	0	0	0	0	0.06	0.12	0	0	0	0

Nd	0	0	0	0	0	0	0.30	0.61	0	0	0	0
Sm	0	0	0	0	0	0	0	0	0	0	0	0
Eu	0	0	0	0	0	0	0	0	0	0	0	0
Gd	0	0	0	0	0	0	0	0	0	0	0	0
Dy	0	0	0	0	0	0	0	0	0	0	0	0
Ho	0	0	0	0	0	0	0	0	0	0	0	0
Er	0	0	0	0	0	0	0	0	0	0	0	0
Tm	0	0	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.01
Yb	0	0	0	0	0	0	0	0	0	0	0	0
Lu	4.09	0.15	4.19	0.70	4.85	0.27	5.64	0.46	7.34	1.26	10.11	2.72
Hf	9.53	2.15	8.31	1.58	7.09	1.52	6.27	0.83	5.52	0.79	5.04	0.40
Ta	0.93	0.22	0.94	0.12	0.71	0.08	0.56	0.09	0.47	0.07	0.36	0.07
W	173.70	13.50	175.82	1.98	182.33	10.07	179.77	5.96	169.24	15.92	159.85	8.85
Re	0	0	0	0	0	0	0	0	0	0	0	0
Pt	0.96	0.11	2.45	0.29	1.84	0.21	1.89	0.31	1.79	0.24	1.72	0.12
Tl	0	0	0.03	0.05	0	0	0	0	0	0	0	0
Pb208	26.45	1.42	13.71	0.45	10.69	1.21	3.22	0.87	0.75	0.34	1.31	0.73
Th	8.13	1.00	6.01	0.57	5.26	0.68	5.13	0.66	4.66	0.36	4.39	0.26
U	0	0	0	0	0	0	0	0	0	0	0	0

B. Elements in the spent medium of 5 l DvH glass reactor 2

	Glass DvH Reactor 2 t0		Glass DvH Reactor 2 t1		Glass DvH Reactor 2 t2		Glass DvH Reactor 2 t3		Glass DvH Reactor 2 t4		Glass DvH Reactor 2 t5	
	nM	**SD**	**nM**	**SD**	**nM**	**SD**	**nM**	**SD**	**nM**	**SD**	**nM**	**SD**
V	10.54	1.94	9.92	0.24	9.83	1.52	9.64	0.86	10.93	1.75	10.25	0.23
Cr53	55.72	13.20	57.73	16.28	43.48	11.69	53.19	3.23	58.17	11.38	42.65	9.06
Mn	19071.82	352.81	17638.64	658.18	17819.32	572.65	16251.14	608.89	18143.86	1603.13	16525.23	566.59
Fe	41631.70	2957.96	35183.04	436.59	29216.52	969.08	13793.30	1143.40	8073.21	1374.46	3581.25	261.05
Co	9179.87	411.11	5435.59	208.98	4047.46	47.20	2259.72	107.58	1559.49	121.24	765.83	282.13
Ni	2738.13	124.96	2445.13	19.67	2348.19	62.74	2093.19	74.20	2143.25	158.16	1609.44	203.02
Cu	922.32	94.36	359.58	49.85	87.80	25.56	32.85	1.09	20.65	5.53	14.45	2.72
Zn66	12561.36	883.95	8928.98	2279.45	3724.43	1612.92	3098.69	1497.55	1992.16	785.19	1606.31	373.59
Zn68	10937.87	759.14	7746.51	2000.07	3222.13	1397.25	2701.82	1296.51	1748.33	698.46	1399.36	332.38
Ga	31.76	4.75	28.15	1.71	22.59	1.88	20.09	1.22	23.03	2.39	20.24	2.20
As	8.85	1.22	8.47	1.58	9.13	0.44	8.02	1.04	7.11	0.86	6.89	1.12
Rb	310.85	16.86	284.21	3.78	298.72	10.57	285.62	11.39	306.49	26.63	286.19	6.14
Sr	278.10	35.33	382.93	274.95	262.24	10.70	238.25	11.74	256.70	23.26	235.40	6.71
Mo95	1516.50	86.61	1052.13	57.46	675.04	71.28	277.58	23.80	191.72	10.65	123.95	45.58
Mo98	1376.14	76.60	946.88	54.01	612.09	65.57	250.18	22.13	172.35	9.37	110.95	38.95
Ru	0	0	0.05	0.04	0.01	0.02	0.00	0.01	0.05	0.05	0	0
Rh	0	0	0	0	0	0	0	0	0	0	0	0
Pd	0	0	1128.68	33.05	915.11	58.19	1206.36	103.83	987.93	170.00	772.50	275.69
Cd111	53.16	1.30	26.84	3.12	5.49	0.38	4.08	1.36	2.49	0.72	0.72	0.47
Sn120	1.53	1.97	1.04	2.02	0.28	0.33	0	0	0	0	0	0
Sb	5.92	0.55	5.58	0.38	3.94	0.33	1.22	0.30	0.14	0.21	0	0
Te	1.31	0.98	2.23	0.33	1.82	0.35	1.55	0.79	1.65	0.25	1.62	0.79
Cs	2.94	2.80	2.92	2.95	1.54	0.12	1.40	0.10	1.15	0.15	1.35	0.34
Ba	83.31	5.89	78.51	6.87	65.33	2.81	63.36	3.50	64.89	1.08	57.52	1.36
La	0.09	0.18	0.37	0.75	0	0	0	0	0	0	0	0
Ce	0.69	0.82	1.45	2.04	0.35	0.25	0.25	0.22	0	0	0	0
Pr	0	0	0	0	0	0	0	0	0	0	0	0
Nd	0	0	0.03	0.06	0	0	0	0	0	0	0	0
Sm	0	0	0	0	0	0	0	0	0	0	0	0
Eu	0	0	0	0	0	0	0	0	0	0	0	0
Gd	0	0	0	0	0	0	0	0	0	0	0	0
Dy	0	0	0	0	0	0	0	0	0	0	0	0
Ho	0	0	0	0	0	0	0	0	0	0	0	0

Er	0	0	0	0	0	0	0	0	0	0	0	0
Tm	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0	0	0.00	0.00
Yb	0	0	0	0	0	0	0	0	0	0	0	0
Lu	5.37	0.20	5.44	0.28	5.18	0.31	5.42	0.37	5.79	0.30	7.13	0.89
Hf	5.38	1.37	4.54	0.53	4.15	0.44	3.87	0.26	4.08	0.22	3.72	0.25
Ta	0.54	0.06	0.52	0.08	0.40	0.02	0.38	0.05	0.29	0.03	0.16	0.03
W	199.06	3.68	193.10	12.78	180.68	4.83	166.01	14.81	157.21	10.56	131.46	6.94
Re	0	0	0	0	0	0	0	0	0	0	0	0
Pt	0.85	0.06	1.18	0.06	1.17	0.07	1.24	0.04	1.37	0.13	1.36	0.20
Tl	0	0	0	0	0	0	0	0	0	0	0	0
Pb208	32.27	3.24	18.79	3.08	6.89	1.25	2.71	0.59	1.14	1.03	0.42	0.36
Th	4.40	0.27	3.84	0.25	3.77	0.24	3.86	0.14	3.66	0.16	3.54	0.20
U	0	0	0	0	0	0	0	0	0	0	0	0

C. Elements in the spent medium of 600 l DvH metal reactor.

	Metal DvH t0		Metal DvH t1		Metal DvH t2		Metal DvH t3		Metal DvH t4		Metal DvH t5		Metal DvH t6	
	nM	SD	nM	SD	nM	SD	nM	SD	nM	SD	nM	SD	nM	SD
V	16.54	1.72	14.17	0.89	16.62	0.74	13.95	1.36	13.12	0.90	13.66	1.83	11.01	1.56
Cr	80.95	4.72	71.50	7.13	82.92	6.29	69.53	7.00	75.81	9.64	72.13	7.12	65.93	18.28
Mn	14903.86	148.29	12891.14	852.27	14247.27	401.03	12863.18	1134.70	12736.36	1605.26	13112.05	1483.22	11738.18	1521.71
Fe	17079.24	254.89	14316.29	161.09	12620.76	258.57	6358.26	1302.07	5354.46	284.44	4076.12	275.41	2087.68	1343.12
Co	7138.35	69.95	1743.50	96.16	1319.43	32.96	1035.89	77.87	370.42	70.96	152.93	25.32	122.54	4.35
Ni	2141.50	20.22	1602.44	197.41	1367.56	40.39	1050.13	32.91	1008.00	64.90	723.38	73.41	815.25	132.09
Cu	78.65	6.10	0	0	0	0	0	0	0	0	0	0	0	0
Zn66	8257.39	159.37	249.38	101.92	918.40	1023.27	506.65	434.27	554.38	295.32	446.42	150.18	216.27	36.78
Zn68	6652.39	133.21	223.95	81.45	762.94	826.79	431.23	353.17	462.77	239.40	374.01	113.59	199.60	35.70
Ga	32.52	1.48	26.97	2.37	30.38	1.89	25.72	3.07	24.82	2.45	25.30	6.04	12.87	0.83
As	6.88	1.14	2.20	0.47	2.66	0.44	1.99	1.23	1.11	0.47	1.56	0.76	1.90	0.83
Rb	98.56	2.22	87.18	5.64	91.20	2.68	82.58	7.54	82.39	5.12	84.72	10.74	74.42	14.00
Sr	116.52	4.72	102.54	4.45	112.78	3.96	100.73	8.62	116.85	17.37	108.19	21.34	104.19	49.67
Mo95	1110.00	11.79	410.84	11.66	344.05	7.30	197.92	40.11	109.60	3.29	93.17	4.56	71.66	4.14
Mo98	1048.47	7.63	384.91	11.88	321.01	5.24	182.55	38.30	102.76	3.94	84.83	3.80	66.26	2.86
Ru	0.04	0.03	0.04	0.06	0.11	0.08	0.18	0.16	0.17	0.03	0.14	0.12	0.39	0.29
Rh	1.17	0.01	1.17	0.02	1.21	0.03	1.20	0.03	1.22	0.02	1.22	0.02	1.38	0.19
Pd	0	0	39.02	6.33	61.74	11.12	0.03	0.06	0	0	8.78	17.56	382.25	573.84
Cd	2.47	0.13	1.23	0.24	1.00	0.37	0.69	0.16	0.68	0.10	0.80	0.14	0.59	0.09
Sn	5.92	0.76	1.12	0.50	1.02	0.35	0.61	0.31	0.03	0.04	0.19	0.35	0.06	0.11
Sb	3.41	0.16	2.22	0.14	2.00	0.16	1.38	0.12	1.08	0.04	0.99	0.02	0.98	0.01
Te	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cs	8.05	0.31	6.47	0.49	6.84	0.42	5.88	0.33	6.29	0.70	6.32	0.60	5.95	0.97
Ba	103.06	4.17	83.80	8.69	91.41	2.99	76.97	5.02	75.83	9.15	73.86	17.65	36.87	4.30
La	0.94	0.11	0.91	0.16	0.97	0.09	0.86	0.08	0.78	0.08	0.95	0.31	0.78	0.11
Ce	0.84	0.19	0.78	0.13	0.85	0.07	0.84	0.10	0.68	0.12	0.94	0.50	0.66	0.10
Pr	1.43	0.01	1.42	0.01	1.43	0.01	1.43	0.01	1.42	0.03	1.46	0.06	1.42	0.03
Nd	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sm	1.65	0.08	1.56	0.04	1.70	0.04	1.62	0.08	1.59	0.08	1.65	0.10	1.60	0.04
Eu	0.44	0.02	0.44	0.02	0.45	0.02	0.45	0.01	0.46	0.02	0.45	0.01	0.45	0.04
Gd	0.96	0.06	0.94	0.05	0.90	0.04	0.90	0.03	0.87	0.06	0.92	0.03	0.83	0.03
Dy	0	0	0	0	0	0	0	0	0.03	0.06	0.14	0.12	0.03	0.05
Ho	0.91	0.01	0.91	0.01	0.92	0.01	0.92	0.01	0.93	0.01	1.00	0.03	0.95	0.02

Er	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Tm	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Yb	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Lu	2.28	0.24	3.29	0.21	3.52	0.19	6.15	2.05	22.06	2.69	45.02	12.39	31.37	3.64
Hf	4.98	0.19	5.19	0.28	5.35	0.29	5.51	0.25	7.07	0.78	9.26	1.11	10.22	2.36
Ta	0.05	0.05	0.13	0.10	0.30	0.05	0.52	0.16	1.42	0.50	2.52	0.22	2.99	1.17
W	243.96	28.70	196.77	16.00	214.78	6.67	191.54	11.25	262.17	44.60	306.06	32.62	199.62	30.96
Re	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Pt	4.77	0.07	5.25	1.09	4.89	0.62	4.60	0.35	4.81	0.47	5.95	0.54	14.26	13.14
Tl	0.31	0.03	0.33	0.02	0.36	0.04	0.36	0.07	0.45	0.05	0.57	0.10	0.43	0.03
Pb	1.07	0.29	0.20	0.24	0.10	0.14	0	0	0	0	0.44	0.54	0.12	0.24
Th	3.54	0.20	3.57	0.12	3.58	0.15	3.72	0.16	4.41	0.31	5.30	0.33	5.56	0.52
U	0.64	0.03	0.60	0.02	0.64	0.02	0.58	0.02	0.64	0.01	0.61	0.03	0.46	0.02

D. Elements in the spent medium of 600 l Pf A11 metal reactor.

	Pf A11 t0		Pf A11 t1		Pf A11 t2		Pf A11 t3		Pf A11 t4		Pf A11 t5		Pf A11 t6	
	nM	SD	nM	SD	nM	SD	nM	SD	nM	SD	nM	SD	nM	SD
V	40.13	1.39	25.54	2.46	31.54	1.34	37.71	1.96	27.48	7.13	30.43	5.31	27.56	5.83
Cr	49.10	7.30	24.99	1.08	30.66	3.45	39.15	11.75	38.49	22.12	36.00	7.31	29.87	4.88
Mn	736.77	24.17	503.66	43.26	576.68	10.33	655.23	20.25	545.11	141.45	618.89	88.45	482.25	132.39
Fe	905.63	90.62	447.92	42.96	525.87	58.82	694.69	230.26	439.52	325.25	304.10	65.93	146.10	82.05
Co	521.00	16.36	321.04	36.17	410.40	8.85	478.86	24.70	363.75	99.12	416.31	79.11	347.29	99.40
Ni	187.20	15.05	127.79	14.52	148.24	4.16	170.76	5.56	133.18	33.39	148.16	22.56	126.27	32.64
Cu	104.92	36.87	56.84	6.78	40.30	3.46	52.17	11.01	70.37	37.38	64.14	12.61	46.13	6.35
Zn66	2886.93	810.65	4134.09	1262.42	2252.27	639.13	1987.33	882.55	2771.93	1315.67	3478.41	1660.02	1521.82	832.36
Zn68	2479.47	690.70	3508.46	1049.67	1948.68	528.71	1730.68	730.94	2388.14	1120.31	2954.28	1374.32	1327.50	710.29
Ga	43.17	1.72	27.11	3.66	35.67	2.42	38.64	2.25	29.10	9.86	36.29	6.95	29.32	9.23
As	2.48	0.88	1.74	0.93	2.25	1.26	1.80	0.53	1.79	1.50	2.07	0.93	1.21	0.51
Rb	140.77	2.70	93.74	8.62	116.24	0.75	133.91	6.53	107.11	35.14	115.42	24.25	92.07	25.33
Sr	162.53	6.42	106.34	10.49	136.12	3.77	155.71	8.39	118.30	28.39	133.23	22.06	107.64	33.08
Mo95	302.64	4.37	199.22	19.56	254.61	3.01	290.84	13.35	225.55	65.86	227.21	43.22	183.32	48.09
Mo98	266.98	6.13	178.09	19.34	225.08	2.64	257.10	12.26	198.10	58.41	202.84	36.99	160.80	43.69
Ru	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Rh	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Pd	0	0	0	0	1.37	2.74	12.05	9.20	24.36	7.95	64.95	9.20	156.62	56.15
Cd	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sn	0.13	0.14	0.28	0.04	0.18	0.18	0.33	0.28	0.93	1.30	0.61	0.33	0.62	0.21
Sb	3.64	0.25	1.56	0.27	2.30	0.20	2.92	0.17	1.85	0.89	2.54	1.20	2.16	1.01
Te	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cs	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ba	143.38	5.04	91.49	11.60	115.18	5.26	131.22	4.28	98.33	28.04	112.32	20.39	88.39	27.26
La	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ce	0	0	0	0	0	0	0	0	0	0	0	0	2.23	4.46
Pr	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Nd	1.02	0.28	0	0	0	0	0	0	0	0	0.14	0.19	0.25	0.50
Sm	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Eu	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Gd	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Dy	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Ho	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Er	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Tm	0.01	0.01	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.01	0.02	0.00
Yb	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Lu	0.02	0.04	0.16	0.30	0.21	0.15	0.14	0.12	0	0	0.04	0.07	0.22	0.29
Hf	7.59	0.07	7.67	0.04	7.74	0.04	7.99	0.07	7.96	0.10	8.64	0.30	8.75	0.18
Ta	0	0	0	0	0	0	0	0	0	0	0.01	0.02	0.16	0.07
W	165.91	3.19	101.50	12.95	127.23	2.10	151.15	4.91	108.43	28.73	119.11	24.22	95.82	27.53
Re	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Pt	1.65	0.07	1.61	0.06	1.70	0.19	1.71	0.07	1.69	0.05	1.75	0.06	1.89	0.10
Tl	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Pb	1.19	0.72	0.36	0.27	0.07	0.07	0.68	0.74	0.62	0.78	0.34	0.24	0.12	0.19
Th	1.93	0.02	1.98	0.04	2.04	0.02	2.08	0.07	2.15	0.05	2.36	0.02	2.59	0.05
U	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 3.S2. The 44 measured elements in the cytoplasmic extracts (S100) and subsequent wash steps given in nM with standard deviations.

A. DvH cytoplasmic extract (S100) and subsequent wash steps.

	DvH S100		DvH S100W		DvH Flow-through		DvH Wash 1		DvH Wash 2		DvH Wash 3	
Metal	Conc. (nM)	**SD**	**Conc. (nM)**	**SD**	**Conc. (nM)**	**SD**	**Conc. (nM)**	**SD**	**Conc. (nM)**	**SD**	**Conc. (nM)**	**SD**
V	0.12	0.14	0.23	0.21	0.11	0.09	0.02	0.04	0.05	0.08	0.06	0.11
Cr53	1.75	1.51	0.70	0.72	1.66	1.60	0.14	0.24	0.40	0.69	0.76	0.81
Mn	640.61	23.54	389.85	6.60	219.59	3.44	71.02	0.80	33.80	0.69	26.12	0.49
Fe	29717.59	1057.05	31960.42	542.25	1134.71	6.59	471.72	23.63	228.48	13.47	252.63	11.10
Co	660.82	25.59	587.35	12.08	51.73	0.76	21.95	0.38	12.65	0.01	10.57	0.03
Ni	156.78	5.21	124.78	17.06	49.90	3.99	19.15	1.16	6.82	0.26	31.67	2.53
Cu	0.00	0.00	2.56	1.07	0.51	0.89	0.00	0.00	0.00	0.00	0.43	0.75
Zn66	0.00	0.00	32.86	35.91	39.10	34.13	58.42	65.44	0.00	0.00	395.56	69.16
Zn68	0.00	0.00	29.99	31.08	33.62	29.27	50.11	56.16	0.00	0.00	334.20	56.66
Ga	2.59	0.15	2.79	0.21	1.61	0.10	3.68	0.27	1.75	0.27	2.05	0.47
As	0.08	0.05	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Rb	5.82	0.31	0.19	0.04	5.62	0.01	1.57	0.14	0.65	0.14	2.22	0.26
Sr	4.21	0.13	1.11	0.12	2.88	0.14	1.40	0.15	0.92	0.16	0.91	0.09
Mo95	71.04	2.01	73.48	2.58	1.58	0.25	0.71	0.10	0.44	0.14	0.68	0.04
Mo98	63.78	1.98	66.25	2.37	1.42	0.16	0.63	0.06	0.36	0.07	0.59	0.04
Ru	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Rh	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pd	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Cd111	0.90	0.05	0.95	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Sn120	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Sb	0.00	0.00	0.00	0.00	0.04	0.04	0.00	0.00	0.00	0.00	0.03	0.02
Te	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

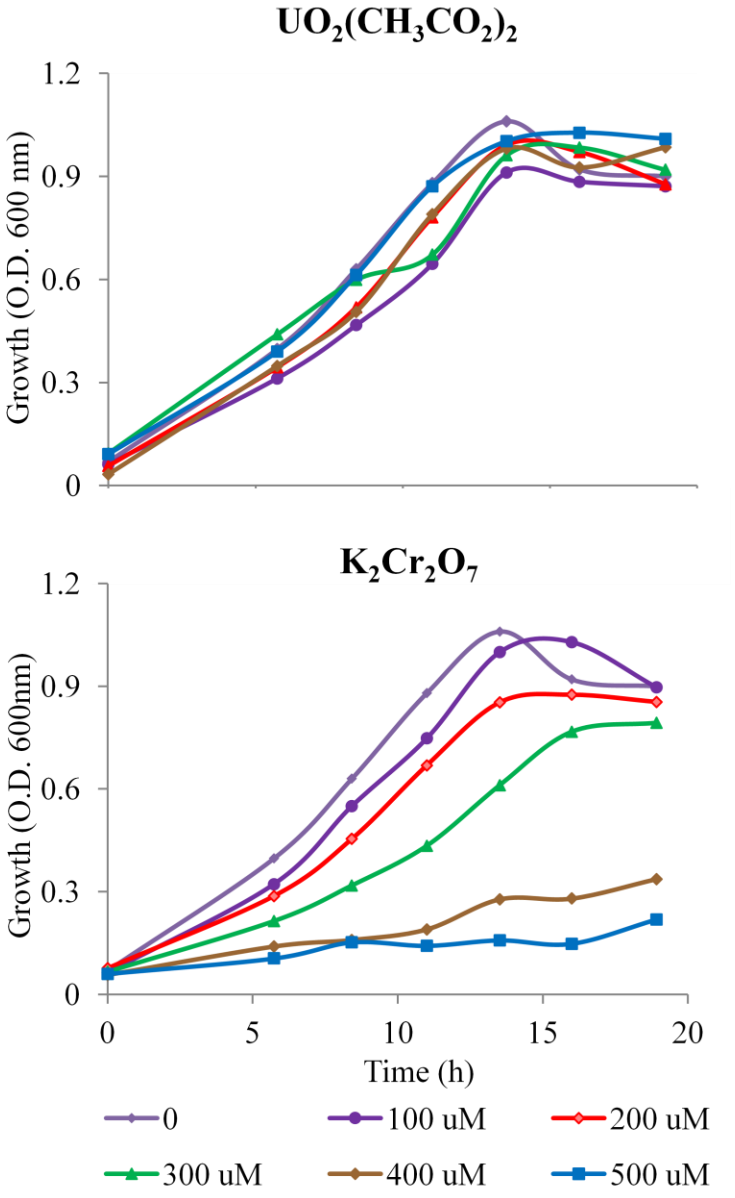
Cs	20.42	1.31	133.33	3.37	220.11	1.28	85.88	1.41	59.13	0.56	1629.78	33.68
Ba	6.60	0.69	6.78	0.64	4.66	0.34	9.91	0.86	4.63	0.50	5.54	0.75
La	0.00	0.00	0.00	0.00	0.04	0.08	0.00	0.00	0.00	0.00	0.00	0.00
Ce	0.00	0.00	0.00	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Pr	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Nd	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Sm	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Eu	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Gd	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Dy	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Ho	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Er	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Tm	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Yb	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Lu	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Hf	0.00	0.00	0.00	0.00	0.03	0.01	0.02	0.01	0.07	0.06	0.04	0.01
Ta	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
W	133.81	5.29	132.31	2.63	1.95	0.02	1.08	0.02	0.59	0.04	0.85	0.02
Re	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pt	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Tl	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pb	0.11	0.11	0.26	0.10	0.03	0.04	0.05	0.05	0.03	0.03	0.10	0.09
Th	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
U	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

B. Pf A11 cytoplasmic extract (S100) and subsequent wash steps.

	Pf A11 S100		Pf A11 S100W		Pf A11 Flow-through		Pf A11 Wash 1		Pf A11 Wash 2		Pf A11 Wash 3	
Metal	**Conc. (nM)**	**SD**	**Conc. (nM)**	**SD**	**Conc. (nM)**	**SD**	**Conc. (nM)**	**SD**	**Conc. (nM)**	**SD**	**Conc. (nM)**	**SD**
V	4.85	0.37	4.46	0.14	0.11	0.11	0.14	0.24	0.11	0.10	0.06	0.06
Cr	17.40	1.62	9.36	1.78	4.01	3.52	1.66	1.47	1.28	1.31	0.69	0.60
Mn	548.52	23.85	433.27	14.04	134.51	2.69	37.92	1.21	17.63	1.25	10.88	0.37
Fe	4931.03	192.00	3844.27	98.33	316.93	25.09	73.66	19.94	38.80	22.29	40.66	11.87
Co	76.68	3.68	69.81	1.98	10.23	0.03	4.34	0.04	2.12	0.08	1.53	0.10
Ni	157.79	13.63	109.41	2.60	60.16	7.62	33.43	7.54	21.38	0.65	26.65	0.79
Cu	49.88	2.10	51.69	1.37	2.54	4.40	0.04	0.06	0.43	0.75	0.00	0.00
Zn66	1121.07	52.35	1454.63	114.15	0.00	0.00	19.53	17.61	46.50	80.54	29.96	51.88
Zn68	947.85	51.71	1231.10	94.03	0.00	0.00	19.29	17.24	40.88	70.81	27.25	47.20
Ga	4.68	0.28	7.45	0.15	6.21	0.28	3.85	1.57	3.02	0.24	3.72	0.53
As	0.13	0.07	0.01	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Rb	34.02	0.37	0.73	0.04	30.60	0.19	9.05	0.25	3.44	0.05	2.10	0.14
Sr	5.12	0.17	2.65	0.32	3.49	0.06	1.61	0.09	2.21	2.83	0.92	0.31
Mo95	166.48	1.82	169.54	3.83	11.91	0.19	6.78	0.25	5.57	0.38	4.40	0.28
Mo98	150.47	1.05	152.97	3.87	10.56	0.11	6.02	0.28	5.01	0.23	4.01	0.16
Ru	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Rh	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pd	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	25.64	10.70
Cd	0.04	0.01	0.12	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Sn	0.43	0.13	0.52	0.08	0.00	0.00	0.05	0.09	0.12	0.20	0.00	0.00
Sb	0.18	0.02	0.06	0.02	0.09	0.01	0.01	0.01	0.00	0.00	0.03	0.04
Te	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Cs	76.32	5.51	226.71	11.56	412.23	14.29	228.19	3.34	227.99	11.15	795.76	6.74
Ba	11.16	1.12	18.84	1.73	15.44	0.86	10.02	4.27	7.29	0.20	10.11	0.63
La	0.04	0.03	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Ce	0.22	0.01	0.30	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Pr	0.02	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Nd	1.76	0.18	2.15	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Sm	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Eu	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Gd	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Dy	0.22	0.02	0.27	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Ho	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Er	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Tm	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Yb	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Lu	0.07	0.06	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Hf	0.07	0.01	0.09	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Ta	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
W	22.65	1.61	11.60	0.67	9.37	0.18	3.38	0.15	1.96	0.09	1.22	0.02
Re	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pt	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Tl	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pb	0.33	0.24	1.02	0.45	0.07	0.06	0.49	0.77	0.05	0.05	0.15	0.12
Th	0.13	0.01	0.09	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U	0.09	0.01	0.05	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Supplementary Figure 3.S1. Growth of Pf A11 grown in media supplemented with uranyl acetate and dichromate (0 – 500 µM). The organism was grown under in the standard medium supplemented as indicated.


$UO_2(CH_3CO_2)_2$
Growth (O.D. 600 nm)
1.2
0.9
0.6
0.3
0
$K_2Cr_2O_7$
Growth (O.D. 600nm)
1.2
0.9
0.6
0.3
0
0
5
10
15
20
Time (h)
0
100 uM
200 uM
300 uM
400 uM
500 uM

CHAPTER 4

THE INFLUENCE OF ENVIRONMENT AND METABOLIC CAPACITY ON THE SIZE OF A MICROORGANISM[3]

[3] Lancaster WA and Adams MWW (2011) p. 93-103. *In* P. L. Luisi and P. Stano (ed.), The Minimal Cell: The Biophysics of Cell Compartment and the Origin of Cell Functionality. Doi:10.1007/978-90-481-9944-0_6.

ABSTRACT

The environment a microorganism inhabits dictates the metabolic capacity necessary for it to survive, and ultimately the minimum size which an organism can achieve. Nutrient rich environments such as those experienced by parasitic bacteria can accommodate organisms with restricted metabolic capacities with relatively few genes, perhaps as few as 250. Nutrient poor environments, such as those experienced by autotrophs provide only minerals and gases and require high biosynthetic capacity to synthesize all cellular carbon from CO_2. This high biosynthetic capacity requires at most 1,500 (an actual value) and perhaps as few as 750 genes. Calculations show that as theoretical minimal cell size is decreased, the cellular volume devoted to the DNA required to encode the minimum gene complement becomes a limiting factor in further reduction. Assuming composition of 50% water, 20% protein, 10% ribosomes and 10% DNA, a spherical cell with minimum biosynthetic capacity (250 genes) would be at least 172 nm in diameter. A cell with high biosynthetic capacity (750 genes) of the same composition would be at least 248 nm in diameter. It is concluded that cells with biochemical requirements for growth, metabolism and reproduction similar to those of known organisms cannot be smaller than 172 nm.

INTRODUCTION

The question of minimum microbial size was dramatically brought to the fore several years ago by the report of McKay and coworkers (19) in which objects with upper dimensions of 20 by 100 nm were postulated to be of cellular origin. These claims have remained controversial more than a decade after their initial publication (25), (11). Nevertheless, the discovery of possible extraterrestrial microbial fossils has spurred interest in ongoing efforts to detect past or

current conditions suitable for life on the surface of Mars. In addition, so-called ultramicrobacteria have been isolated from marine environments that can pass through a 200 nm filter and have cell volumes of 0.03 to 0.08 μm^3(5). Recently one of these ultramicrobacteria, *Herminiimonas glaciei*,was isolated from an ice core taken at a depth of three kilometers from a Greenland glacier (16). The organism was successfully grown in laboratory culture after lying dormant for approximately 120,000 years. Entities known as "nanobacteria" having diameters as low as 80 nm and implicated in human disease were reported to have been cultured from blood (13), which suggested a very low limit on microbial cell size. Subsequent study has shown that these entities are most likely calcium carbonate precipitates (17), leaving the question of the lower limit on microbial cell size very much open. In light of these and other studies, it is important to estimate a reasonable limit for the theoretical minimum size of microorganisms. Would it be possible for a cell with 50 nm diameter to contain sufficient biological material to sustain a free-living lifestyle? This immediately raises the question of what minimum quantity of biological material is sufficient. One measure that has been increasingly informative as the quantity of sequence data has exploded is genome size, which determines diversity of proteins (enzymes) that an organism has at its disposal to support growth, metabolism and reproduction.

The minimum complement of genes necessary to maintain life is intensely dependant on the biosynthetic capacity of a given organism, which is in turn largely determined by the environment in which the organism is found. It is no surprise that a wide range of biosynthetic capacities exists among organisms given the extent to which dependence on the environment varies. All organisms must satisfy the same basic biochemical requirements, but some environments provide only basic compounds necessary for life as we know it, while other provide a variety of complex molecules, obviating the need for their synthesis by an organism

inhabiting such an environment. Minimum requirements for human survival include ten or so amino acids, various minerals, a variety of biological cofactors (vitamins), and a continual supply of oxygen. Like us, the vast majority of microorganisms require a fixed carbon source, which is usually a carbohydrate although some can utilize lipids, nucleotides, or various simple organic compounds. Microbes, however, differ from humans in their ability to synthesize most of the amino acids as well as essential cofactors which we must obtain from our environment. In contrast, some microorganisms are intensely dependent upon their environments. For example, some microbial parasites do not synthesize any amino acid or lipid, and only a few vitamins and nucleotides; rather, they obtain all of these compounds from their host. One interesting class of organisms completely dependent on a host organism is the mutualistic endosymbionts, one example of which is *Carsonella rudii* which provides essential amino acids to their insect hosts and lack genes for many essential synthetic pathways. The genome of *Carsonella rudii* is only 160 kilobases, which encodes 182 open reading frames, making it the smallest sequenced bacterial genome (21). The extreme reduction of the *C. rudii* genome and the degree of its interaction with the host have led to the proposal that it represents a transition from a symbiotic cell to an organelle of the host cell (24).

Such reduced life-forms approach the environmental dependence of a virus, which consists of a protein coat that surrounds a defined amount of nucleic acid (DNA or RNA). It has recently been proposed that biological entities should be grouped into two major categories, Ribosome-encoding organisms (REOs) and Capsid-encoding organisms (CEOs) (22). In this scheme, REOs include free living cells, parasites, endosymbionts and organelles, CEOs include viruses, with plasmids and viroids considered “orphans”, encoding neither ribosomes nor capsids. The CEOs rely on their host organism not only for basic biological compounds, but for

the transcriptional and translational machinery of REOs necessary for their own reproduction. Along with the plasmids and viroids, these entities are able to achieve very small size by encoding minimal information necessary for their replication and exploiting the host environment to provide the complex machinery and metabolic activity necessary to support it.

Plasmids, Viruses and parasitic Bacteria obviously depend greatly on their environments to provide complex and varied components necessary for their survival. Other organisms however, thrive in quite sparse environments. Such organisms require only simple chemicals, such as CO_2, O_2, H_2, and NH_3 and synthesize all amino acids, cofactors, nucleotides, etc., with CO_2 as the sole carbon source, and ammonia (or even N_2 gas) as the nitrogen source. These autotrophs may obtain the energy used to synthesize these compounds from light, as in green plants and cyanobacteria, or from the oxidation of abundant compounds present in their environments. Substances oxidized by aerobic autotrophs include H_2, CO, CH_4, NH_3, or H_2S. Similarly, anaerobic autotrophs growing on H_2 and CO_2 also conserve energy during the reduction of CO_2, either by the production of methane or acetate as accomplished by methanogens and acetogens, respectively. There is obviously a spectrum of environmental dependence with a rich diversity of organisms represented at the aforementioned extremes and everywhere in between.

Given the vast differences in the dependence of organisms on their environments, it should be clarified what is meant by "free-living" in the context of elucidating the minimal hypothetical cell. All life forms must at a minimum undergo self-directed reproduction when supplied with the molecules they need in sufficient quantities and, in some cases, non-molecular sources of free energy in the form of electromagnetic radiation. This minimum requirement is met by entities which do not construct the machinery necessary for the execution of this

replication. The environment of a CEO, a cell with an array of replicative machinery which can be usurped, clearly allows these entities to attain a very small size indeed while allowing replication. In fact, an even greater reduction is possible for plasmids which rely on the host cell not only for the copying of genetic information, but with its maintenance and transmission. There is a major distinction to be made between such replicators and even the most diminutive parasitic bacterium which, while relying on an array of complex molecules from the environment, does not reproduce via a transfer of genetic material into the environment. Although parasitic cellular life may depend on the complex substances provided by a living host, these nutrients, if provided to the organism, replace the host and allow the parasite to grow and reproduce. This is not the case with viruses and plasmids which rely on the intact functioning system of the host. These entities will not therefore be considered in determining the lower size limit for life.

As has been discussed, life forms occupy a wide range of environments which greatly affect their metabolic capacities. In determining the lower limit on microbial cell size, it will be useful to consider two extremes, the rich environment experienced by parasitic life forms, and the sparse environment providing only gasses and minerals experienced by many autotrophs. In determining these limits, it is vital to estimate the number of proteins required by our two classes of organisms, those with high and low biosynthetic capacities. The ever increasing availability of genome sequences for a variety of microorganisms enables quantitative estimations to be made.

ORGANISMS WITH LOW BIOSYNTHETIC CAPACITY

Those organisms that are most dependent upon their environment are the parasitic bacteria, the prototypical example of which are the *Mycoplasma*. The genome of *Mycoplasma genitalium* was the second to be completely sequenced (8). At 0.58 Mb, this still represents the smallest known genome of any free-living organism. The genome contains 470 predicted protein coding regions, and these include those required for DNA replication, transcription and translation, DNA repair, cellular transport, and energy metabolism. An early comparative genomic study utilizing the genome of the first organism to be completely sequenced, *Haemophilus influenzae* (6), led to the conclusion that the "minimal gene set that is necessary and sufficient to sustain the existence of a modern-type cell" is (only) 256 genes, or about half of the genome of *M. genitalium* (20). Subsequent comparative genomic and experimental studies have yielded a range of estimates for the minimal genome complement (3) and the concept of what constitutes a minimal gene set has evolved as greater amounts of sequence data have become available. A recent study involving 573 bacterial genomes identified an "extended core" set of ~250 protein families present in almost all species (14) which may be considered as necessary but not necessarily sufficient for a modern minimal bacterial cell. It should be noted, however, that while some parasitic organisms such as *Mycoplasma* species grow in the absence of their hosts, to do so they require an extremely rich medium which mimics the availability of nutrients encountered in the infected host. These organisms maintain a minimal biosynthetic capacity, a capacity that could reasonably be satisfied by approximately 250 different proteins.

THE MOST SLOWLY EVOLVING MICROORGANISMS

In determining the "minimum" set of genes that a minimal-size microbe might contain, we must also consider what is meant by the term "modern-type cell" quoted above (20). Are all present-day organisms highly sophisticated cells with a range of metabolic capabilities, only some of which are utilized and then under very specialized conditions? For example, *E. coli* could be regarded as highly evolved because it exhibits a range of metabolic modes, including growth under aerobic and anaerobic conditions, the utilization of a wide variety of different carbon sources, etc. Indeed, such a large metabolic capacity might be reflected in its genetic content of 4.64 Mb encoding 4,288 genes (1). Similarly, metabolically diverse species such as *Bacillus subtilus* and *Pseudomonas putida* have genomes of comparable size (≥4 Mb). The average genome length of 1,113 sequenced bacterial genomes including large genomes such as *Sorangium cellulosum* (13Mb) is 3.6 Mb, (15). In other words, it is not unusual for well characterized microorganisms to contain 4,000 or more genes, in some cases many more. Hence, are there more-slowly-evolving organisms, and do they contain less genetic material and have fewer metabolic choices?

By phylogenetic analyses based on 16S rRNA sequence comparisons, the most-slowly-evolving microorganisms are the deepest branches, the first to have diverged within the two major lineages corresponding to the Bacteria and the Archaea (27), (23). Remarkably, in both domains these are the hyperthermophiles, organisms that grow optimally at temperatures of 80°C and above. Within the bacteria domain this includes few genera, the most well studied of which are *Thermotoga* and *Aquifex*, while there are more than 20 genera of hyperthermophilic archaea (15). In fact, one of the two major branches within the archaeal domain consists almost entirely of hyperthermophiles, while in the other branch the hyperthermophiles are the most slowly

evolving of the known genera. A great deal is known about the genome contents of these hyperthermophilic organisms as representatives were among the first sequenced genomes, and there is continuing interest in the potential applications of thermostable enzymes. There are currently 31 completely sequenced hyperthermophilic archaeal and seven hyperthermophilic bacterial genomes, with many more ongoing sequencing projects (15). Interestingly, most of these organisms have genomes only about half the size of that found in *E. coli*, with those of *Sulfolobus solfataricus* and *Ignicoccus hospitalis* being the largest (2.99 Mb) and smallest (1.30 Mb) of the free-living members of this group, respectively with an average of 2.04Mb for all the hyperthermophilic organisms in this set. Thus, the most slowly evolving organisms known (at least as determined by 16S rRNA analyses) do indeed have relatively small genomes, although they are still highly complex life-forms. Interestingly, *Ignicoccus hospitalis*, the free-living hyperthermophile with the smallest sequenced genome is host to the smallest hyperthermophilic parasite, *Nanoarchaeum equitans* (12). *N. equitans* has a diameter of only 400 nm, and its 491 kilobase genome containing 552 predicted open reading frames (26) is very close in size to that of *M. genitalium*. Experiments have shown that organisms closely related to the sequenced *Ignicoccus* and *Nanoarchaeum* species are among the first to colonize nascent hydrothermal vents (18). Nanoarchaeal sequences have been identified in a wide range of environments, but whether all members are obligate parasites, or there are free-living representatives remains unknown (7).

ORGANISMS WITH HIGH BIOSYNTHETIC CAPACITY

The genome sequences of autotrophs can provide us with information on the upper limit of the number of different protein encoding genes necessary for organisms living in sparse environments providing only gases and minerals. To date, the genomes of several hyperthermophilic autotrophs have been sequenced. The first to be sequenced was the archaeon, *Methanococcus jannaschii*, (later reclassified as *Methanocaldococcus jannaschii)*, (2), which is a methanogen that grows up to 90°C using H_2 and CO_2 as energy and carbon sources and generates methane as an end product. The genome of *Methanocaldococcus jannaschii* is 1.7Mb with approximately 1,700 protein encoding genes. The average genome length of eight sequenced hyperthermophilic autotrophs, seven of which are archaea, is 1.7Mb(15). This group includes anaerobic sulphur , nitrate and arseneic reducers and methanogens. The one bacterium in this group is the microaerophilic *Aquifex aeolicus* (4), which grows up to 95°C on H_2 and CO_2. The genome of *Aquifex aeolicus* is 1.5 Mb with approximately 1,500 protein encoding genes. It should be noted that the pathway of CO_2 assimilation and the biochemistry of energy conservation across these sequenced examples is quite varied, yet their metabolic requirements are met by genomes of similar size. On the other hand, these genomes are much larger than those of the parasitic organisms discussed above. Presumably, these organisms require many more genes because they must synthesize all of their cellular components from CO_2. Hence they contain about three times the genetic information of *M. genitalium* This seems appropriate considering that the latter organism is supplied with all of its amino acids, nucleotides, fatty acids, "vitamins," and an energy source (glucose). From this direct comparison we might conclude that about two-thirds of the DNA in *A. aeolicus* and *M. jannaschii*, or approximately

1,000 genes, encodes proteins that function to carry out these biosynthetic tasks and produce all of these compounds from CO_2.

THE SMALLEST CELL

From the above discussion it can be concluded that a cell with minimal biosynthetic capacity that is growing in a nutrient-rich medium requires between 250 (a calculated value) and 500 genes (the approximate number in M. genitalium) to grow. At the other extreme is a cell that synthesizes all of its cellular material from CO_2, and this requires at most 1,500 genes (the approximate number in *A. aeolicus* and *I. hospitalis*) and probably closer to 750 genes (half of the actual value). With these values in mind, it is instructive to consider the amount of biological material which can be contained within a hypothetical small cell of 50 nm diameter. Assuming a 5 nm thick lipid bilayer, a spherical cell of 50 nm diameter would have an internal volume of 33,500 nm^3. An *E. coli* cell, with dimensions of about 1.3 by 4.0 µm, has an internal volume of about 5×10^9 nm^3, or almost two million times the volume of the 50 nm diameter cell. What quantities of the various biochemical structures found in a typical prokaryotic cell can be accommodated within this hypothetical cell with six orders of magnitude less space than well characterized real cells? A ribosome has a diameter of about 20 nm, and ribosomes are typically 25% of the mass (dry weight) of a bacterial cell (although this varies considerably depending on the growth rate). Assuming a similar percentage of the volume of a 50 nm diameter cell is devoted to them, such a cell could contain only two ribosomes (4,200 nm^3 each). Whether only two would limit cell growth to any extent is unknown; nevertheless, the cell is certainly large enough to contain ribosomes, albeit only two. On the other hand, proteins usually constitute about half of the dry weight of bacterial cells. Let us assume that they also occupy approximately

50% of the volume of the 50 nm diameter cell, and that, in general, proteins have an average molecular weight of 30 kDa, which corresponds to a diameter of about 4 nm per protein. If a cell of 50 nm diameter were 50% protein by volume, then this would correspond to about 520 such molecules (average 30 kDa) per cell.

Are two ribosomes and 520 "average-sized" protein molecules sufficient to support the growth of a cell? This amount of protein represents on average, only two copies of each protein for a cell with minimal biosynthetic capacity (predicted to contain 250 genes). Next, the amount of DNA such a cell could contain must be considered. DNA typically occupies only about 3% of the total mass (dry weight) of well characterized bacterial cells, so upon first consideration, it would seem that the volume of genetic material especially in an organism with a minimum gene content, would be a major limiting factor on cellular volume. For example, with a diameter of 2 nm and length of 0.34 nm/bp, the 4.64 Mb of E. coli has a volume of 4.9×10^6 nm^3, which is less than 1% of the cell volume. The 0.49 Mb *N. equitans* genome comprises only about 1.5% of the volume of these 400 nm diameter cells. Surprisingly, however, simple calculations show that DNA is a determining factor in much smaller cells. Thus, the hypothetical 50 nm diameter cell, 75% of which (by volume) is occupied by proteins and ribosomes, could contain, even if the remaining 25% of the cell were devoted to DNA, only eight genes (of 1,000 bp each)!

DNA CONTENT DETERMINES CELL SIZE

Since our hypothetical 50 nm cell has revealed the major volume constraint imposed by DNA, it is clear that the minimum necessary amount of genetic material can serve as a basis for computing a minimum cell size. Even if a hypothetical cell could devote half of its total internal volume to DNA, a cell with minmal biosynthetic capacity would need to have a diameter of 110

nm to contain its 250 genes. This increase in volume also provides space for additional components other than DNA. Assuming that half of the remaining volume (25%) of the 110 nm diameter cell is devoted to protein, and 12.5% to ribosomes, the cell could contain 4,000 kDa proteins (eight per gene) and 15 ribosomes. A hypothetical minimal autotroph with 750 genes and the same proportion of components (50% DNA, 25% protein, 12.5 % ribosomes) would have a diameter of 156 nm. This 156 nm cell could contain 12,400 (16 per gene) proteins and 48 ribosomes. Cells with this proportion of components would have 12.5% of their volume remaining for other cellular components such as lipids, metabolites, cofactors and inorganic substances. These minimum volumes have neglected one vital component, water, which typically occupies the greatest volume of microbial cells. If it is assumed that the hypothetical minimum cell is comprised of 50% water, the minimum computed volumes rise to 136 and 194 nm for cells with minimal and maximal biosynthetic capacity, respectively.

It is clear from these calculations that DNA represents a critical factor in the minimal limit on cell size. The calculations made the unrealistic assumption that 25% of the cell's volume could be devoted to DNA. If a more accurate estimate of minimal cell size is to be reached, a more reasonable figure for the amount of DNA which can be contained in a cell should be used. As previously mentioned, E. coli DNA occupies less than 1% of cellular volume, and it does not seem likely that more than 10% of the volume of any viable functioning cell could be devoted to the chromosome. While there are additional serious constraints on the packing of DNA such as strand bending and negative charge repulsion and the necessity of unwinding the strand for transcription and replication which may further limit the maximum proportion, assuming a content of 10% DNA should yield an informative estimate of a reasonable lower limit on cell size. Thus, a cell that contains 250 genes that occupy 10% of its volume would be 172 nm in

diameter (Figure 1b), while one containing 750 genes occupying the same relative volume would be 248 nm in diameter(Figure 1c). Assuming such cells contain 20% by volume protein and 10% by volume ribosomes, the 172 nm cell could accommodate 64 ribosomes and over 16,000 proteins, or 65 per gene, and the 248 nm cell could contain three times as much.

It may be concluded, therefore, that the minimum theoretical size for a cell is 172 nm in diameter. Since such a cell has a minimal biosynthetic capacity and a reduced genome of 250 genes, it would need to be supplied with all amino acids, fatty acids, nucleotides, cofactors, and other components to survive. The cell would have no cell wall.and would consist, by volume, of 10% DNA, 10% ribosomes, 20% protein, and 50% water. It would contain approximately 65 proteins per gene as well as 65 ribosomes. In contrast, an autotrophic cell with a much higher biosynthetic capacity, such that it could synthesize all cellular components from CO_2, would be 248 nm in diameter, assuming that its DNA is limited to 10% of the cell volume. Note that these calculations assume a theoretical minimum gene content, which is about half of that present in known life-forms. The amount of DNA in a known autotrophic organism (approximately 1,500 genes in *A. aeolicus*) would require a cell of at least 314 nm in diameter, assuming that it occupied 10% of the cell by volume. Hence, depending on the biosynthetic capacity of a cell, and the extent to which the calculated minimum gene content is realistic, its minimum diameter is between 172 and 314 nm. Overall, one can conclude that microorganisms cannot have diameters significantly less than 200 nm if they have the same basic biochemical requirements for growth as all other extant life-forms, but even then they would require very specialized environments. More likely, the absolute minimum size is closer to 250 nm where the cell is able to grow on simple compounds commonly found in various natural environments.

To date, no free-living organism close to the hypothetical 250 nm minimum diameter has been identified. What place would such a minimal organism have in the modern world? In natural environments, an organism must cope with a myriad of challenges such as constantly changing metabolite concentration, changing temperature, the intrusion of toxic compounds and radiation, as well as competition from other organisms. The ability to deal effectively with such situations will undoubtedly outweigh the metabolic cost of maintaining some number of genes in addition to the minimal core. This may not have been the case in archaic ecosystems with lower diversity of cellular life and viruses, as well as potentially reduced replicative fidelity relative to that present in modern systems. Whether or not ancestral microbes resembled our concept of minimal life or not, there is a place for minimal life in the modern world. The field of synthetic biology is burgeoning, and attempts are being made to construct a minimal cell (10), (9) both as a tractable perturbable system in which to study the fundamentals of life, and as a platform for creating custom organisms with known behavior to produce useful substances and perform specific tasks. In the latter regard, the minimal cell has two primary advantages. The first is that such a system can potentially achieve greater efficiency and more predictable behavior by adding only that which is necessary to perform the desired task relative to engineering a system based on a versatile organism such as *E. coli*. The other is that the metabolic and environmental fragility becomes a distinct advantage in certain cases. In producing biological substances in a laboratory environment, stable conditions can be maintained, and the real or perceived danger of engineered microbes escaping into natural ecosystems causing damage is reduced. The utility of such organisms built from the "ground up" remains to be seen, but the potential to further our understanding of fundamental life processes as well as perform useful tasks and produce useful compounds is staggering.

ACKNOWLEDGMENTS

Research carried out in the authors' laboratory was funded by the US National Science Foundation and the US Department of Energy

REFERENCES

1. **Blattner, F. R., G. Plunkett, III, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao.** 1997. The Complete Genome Sequence of Escherichia coli K-12. Science **277:**1453-1462.

2. **Bult, C. J., O. White, G. J. Olsen, L. Zhou, R. D. Fleischmann, G. G. Sutton, J. A. Blake, L. M. FitzGerald, R. A. Clayton, J. D. Gocayne, A. R. Kerlavage, B. A. Dougherty, J.-F. Tomb, M. D. Adams, C. I. Reich, R. Overbeek, E. F. Kirkness, K. G. Weinstock, J. M. Merrick, A. Glodek, J. L. Scott, N. S. M. Geoghagen, J. F. Weidman, J. L. Fuhrmann, D. Nguyen, T. R. Utterback, J. M. Kelley, J. D. Peterson, P. W. Sadow, M. C. Hanna, M. D. Cotton, K. M. Roberts, M. A. Hurst, B. P. Kaine, M. Borodovsky, H.-P. Klenk, C. M. Fraser, H. O. Smith, C. R. Woese, and J. C. Venter.** 1996. Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus jannaschii. Science **273:**1058-1073.

3. **Carbone, A.** 2006. Computational prediction of genomic functional cores specific to different microbes. J Mol Evol **63:**733-46.

4. **Deckert, G., P. V. Warren, T. Gaasterland, W. G. Young, A. L. Lenox, D. E. Graham, R. Overbeek, M. A. Snead, M. Keller, M. Aujay, R. Huber, R. A. Feldman, J. M. Short, G. J. Olsen, and R. V. Swanson.** 1998. The complete genome of the hyperthermophilic bacterium Aquifex aeolicus. Nature **392:**353-8.

5. **Eguchi, M., T. Nishikawa, K. Macdonald, R. Cavicchioli, J. C. Gottschal, and S. Kjelleberg.** 1996. Responses to Stress and Nutrient Availability by the Marine Ultramicrobacterium Sphingomonas sp. Strain RB2256. Appl Environ Microbiol **62:**1287-1294.

6. **Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J. F. Tomb, B. A. Dougherty, J. M. Merrick, and et al.** 1995. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science **269:**496-512.

7. **Forterre, P., S. Gribaldo, and C. Brochier-Armanet.** 2009. Happy together: genomic insights into the unique Nanoarchaeum/Ignicoccus association. J Biol **8:**7.

8. **Fraser, C. M., J. D. Gocayne, O. White, M. D. Adams, R. A. Clayton, R. D. Fleischmann, C. J. Bult, A. R. Kerlavage, G. Sutton, J. M. Kelley, R. D. Fritchman, J. F. Weidman, K. V. Small, M. Sandusky, J. Fuhrmann, D. Nguyen, T. R.**

Utterback, D. M. Saudek, C. A. Phillips, J. M. Merrick, J. F. Tomb, B. A. Dougherty, K. F. Bott, P. C. Hu, T. S. Lucier, S. N. Peterson, H. O. Smith, C. A. Hutchison, 3rd, and J. C. Venter. 1995. The minimal gene complement of Mycoplasma genitalium. Science **270:**397-403.

9. **Gibson, D. G., G. A. Benders, C. Andrews-Pfannkoch, E. A. Denisova, H. Baden-Tillson, J. Zaveri, T. B. Stockwell, A. Brownley, D. W. Thomas, M. A. Algire, C. Merryman, L. Young, V. N. Noskov, J. I. Glass, J. C. Venter, C. A. Hutchison, 3rd, and H. O. Smith.** 2008. Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. Science **319:**1215-20.

10. **Glass, J. I., N. Assad-Garcia, N. Alperovich, S. Yooseph, M. R. Lewis, M. Maruf, C. A. Hutchison, 3rd, H. O. Smith, and J. C. Venter.** 2006. Essential genes of a minimal bacterium. Proc Natl Acad Sci U S A **103:**425-30.

11. **Golden, D. C., D. W. Ming, R. V. Morris, A. J. Brearley, H. V. Lauer, Jr., A. H. Treiman, M. E. Zolensky, C. S. Schwandt, G. E. Lofgren, and G. A. McKay.** 2004. Evidence for exclusively inorganic formation of magnetite in Martian meteorite ALH84001. American Mineralogist **89:**681-695.

12. **Huber, H., M. J. Hohn, R. Rachel, T. Fuchs, V. C. Wimmer, and K. O. Stetter.** 2002. A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. Nature **417:**63-7.

13. **Kajander, E. O., I. Kuronen, K. K. Akerman, A. Pelttari, and N. Ciftcioglu.** 1997. Presented at the Instruments, Methods, and Missions for the Investigation of Extraterrestrial Microorganisms, San Diego, CA, USA.

14. **Lapierre, P., and J. P. Gogarten.** 2009. Estimating the size of the bacterial pan-genome. Trends Genet **25:**107-10.

15. **Liolios, K., K. Mavromatis, N. Tavernarakis, and N. C. Kyrpides.** 2008. The Genomes On Line Database (GOLD) in 2007: status of genomic and metagenomic projects and their associated metadata. Nucleic Acids Res **36:**D475-9.

16. **Loveland-Curtze, J., V. I. Miteva, and J. E. Brenchley.** 2009. Herminiimonas glaciei sp. nov., a novel ultramicrobacterium from 3042 m deep Greenland glacial ice. Int J Syst Evol Microbiol **59:**1272-7.

17. **Martel, J., and J. D. Young.** 2008. Purported nanobacteria in human blood as calcium carbonate nanoparticles. Proc Natl Acad Sci U S A **105:**5549-54.

18. **McCliment, E. A., K. M. Voglesonger, P. A. O'Day, E. E. Dunn, J. R. Holloway, and S. C. Cary.** 2006. Colonization of nascent, deep-sea hydrothermal vents by a novel Archaeal and Nanoarchaeal assemblage. Environ Microbiol **8:**114-25.

19. **McKay, D. S., E. K. Gibson, Jr., K. L. Thomas-Keprta, H. Vali, C. S. Romanek, S. J. Clemett, X. D. Chillier, C. R. Maechling, and R. N. Zare.** 1996. Search for past life on Mars: possible relic biogenic activity in martian meteorite ALH84001. Science **273:**924-30.

20. **Mushegian, A. R., and E. V. Koonin.** 1996. A minimal gene set for cellular life derived by comparison of complete bacterial genomes. Proc Natl Acad Sci U S A **93:**10268-73.

21. **Nakabachi, A., A. Yamashita, H. Toh, H. Ishikawa, H. E. Dunbar, N. A. Moran, and M. Hattori.** 2006. The 160-kilobase genome of the bacterial endosymbiont Carsonella. Science **314:**267.

22. **Raoult, D., and P. Forterre.** 2008. Redefining viruses: lessons from Mimivirus. Nat Rev Microbiol **6:**315-9.

23. **Stetter, K. O.** 2006. Hyperthermophiles in the history of life. Philos Trans R Soc Lond B Biol Sci **361:**1837-42; discussion 1842-3.

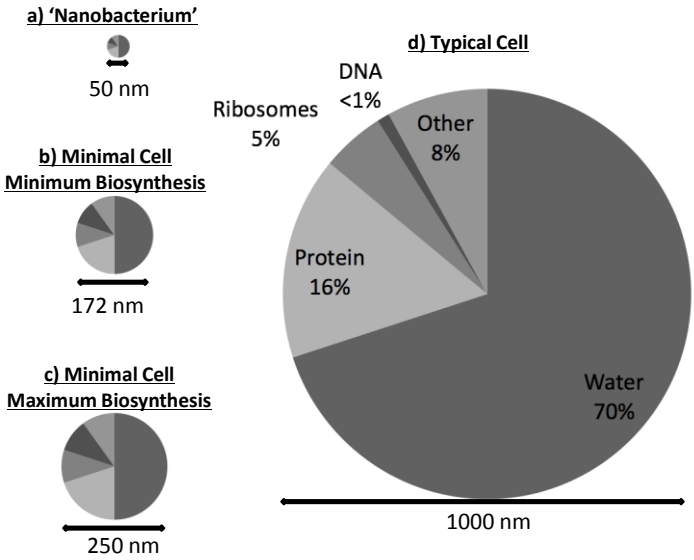
24. **Tamames, J., R. Gil, A. Latorre, J. Pereto, F. J. Silva, and A. Moya.** 2007. The frontier between cell and organelle: genome analysis of Candidatus Carsonella ruddii. BMC Evol Biol **7:**181.

25. **Thomas-Keprta, K. L., S. J. Clemett, D. A. Bazylinski, J. L. Kirschvink, D. S. McKay, S. J. Wentworth, H. Vali, E. K. Gibson Jr, Jr., and C. S. Romanek.** 2002. Magnetofossils from ancient Mars: a robust biosignature in the martian meteorite ALH84001. Appl Environ Microbiol **68:**3663-72.

26. **Waters, E., M. J. Hohn, I. Ahel, D. E. Graham, M. D. Adams, M. Barnstead, K. Y. Beeson, L. Bibbs, R. Bolanos, M. Keller, K. Kretz, X. Lin, E. Mathur, J. Ni, M. Podar, T. Richardson, G. G. Sutton, M. Simon, D. Soll, K. O. Stetter, J. M. Short, and M. Noordewier.** 2003. The genome of Nanoarchaeum equitans: insights into early archaeal evolution and derived parasitism. Proc Natl Acad Sci U S A **100:**12984-8.

27. **Woese, C. R., O. Kandler, and M. L. Wheelis.** 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A **87:**4576-9.

Figure 4.1. A comparison of the diameters of theoretical minimal cells. A comparison of the diameters (drawn to scale) and compositions of (a) a 50 nm hypo-thetical ‘nanobacterium’, (b) a minimal cell with minimal biosynthetic capacity, (c) a minimal cell with maximal biosynthetic capacity, and (d) a typical 1 mm prokaryotic cell with a composition based on that reported for E. coli, which includes ≤1% of the total volume donated to DNA (Neidhardt 1996). The minimal cell sizes (a–c) are based on compositions of 10% DNA, 50% water, 20% protein, 10% ribosomes and 10% other components.


a) 'Nanobacterium'
50 nm
b) Minimal Cell
Minimum Biosynthesis
172 nm
c) Minimal Cell
Maximum Biosynthesis
250 nm
d) Typical Cell
DNA
<1%
Ribosomes
5%
Other
8%
Protein
16%
Water
70%
1000 nm

CHAPTER 5

DISCUSSION

This research has made use of genomic, metallomic and proteomic methods to make progress towards defining the genome and metalloproteome of three model organisms, *Pyrococcus furiosus*, *Desulfovibrio vulgaris* strain Hildenborough, and *Pelosinus fermentans* strain A11. Significant insights were gained into several aspects of the biology of these organisms. A comprehensive chromatography-based metalloproteomic investigation of the soluble proteome of *P. furiosus* identified 343 metal peaks in second column-level fractions, of which 158 were found to have no corresponding known or predicted metalloproteins (4). This indicated a large number of unidentified metalloproteins present in this organism. The processing and analysis of the data set generated in this investigation yielded predictions of likely novel metalloprotein candidates (15). The generation of a complete genome sequence for a naturally competent strain of *P. furiosus* significantly contributes to future genetic engineering efforts which will greatly simplify the verification of these predictions through purification and characterization of candidate proteins (2). In addition, similar metallomics investigations were carried out on two additional organisms resulting in the discovery of some unique aspects of metal assimilation, including significant uptake of uranium by *P. fermentans* A11, which expands our understanding of microbial-metal interaction in contaminated environments and will aid efforts to harness them for cost-effective remediation. In addition, the results of the research presented herein provide clear direction to future areas of research that will expand the breadth and depth of our knowledge of these organisms.

MINIMAL CELLS

Chapter 2 discussed the minimal genetic requirements of organisms with various lifestyles, and the size constraints placed on such organisms. The interest in minimal cells has increased dramatically with the success of efforts to create a viable organism with a chemically synthesized genome (8). The 1.08 Mb chromosome of *Mycoplasma mycoides* JCVI-syn1.0, nicknamed "Synthia" was based on *M. mycoides* subspecies capri. In the course of creating viable self-replicating cells using this genome, it was discovered that a single base pair deletion sequencing error had occurred in the essential dnaA gene, which delayed progress significantly (8). This further underscores the predictive improvements which will be likely as the quantity and quality of genome sequence increases. Although the genome of this organism has only limited modifications from the wild-type organism on which it is based, the methodology will enable future progress in the bottom-up creation of synthetic organisms (17). The announcement of the creation of "synthetic life" created significant controversy into the ethical implications, and whether such technology should be regulated. Significant public discussion ensued and The Presidential Commission for the Study of Bioethical Issues issued a report that recommended 5 guiding principles for evaluating such new technology: public beneficence, responsible stewardship, intellectual freedom and responsibility, democratic deliberation, and justice and fairness (10). Commentaries on the potential benefits and dangers of synthetic biology have continued and will likely continue as the field develops (5, 7).

METALLOPROTEINS PREDICTION IN *P. FURIOSUS*

A metalloproteomic investigation of the soluble proteome of *P. furiosus* yielded 2,589 chromatographic fractions in which 870 total proteins were identified (4, 15). The data set was used to predict novel metalloproteins for 10 metals, cobalt, iron, manganese, molybdenum, nickel, lead, tungsten, uranium, vanadium, and zinc. The likelihood of metal-protein association was determined by the co-occurrence of a protein with a metal throughout the fractionation. Clustering of potential metal associated proteins was used to estimate the number of metalloproteins which explain the observed metal peaks. This ranged from 5 predicted novel vanadium proteins to 13 for cobalt. One or more proteins occurring in each cluster represent a novel metalloprotein or protein complex. Since the results of this research were published, an important new tool to experimentally verify predicted metalloprotein candidates has become available, a naturally competent strain of *P. furiosus* (16). As mentioned previously, determining the biologically correct metal using heterologous expression has serious limitations. The use of a native expression system including the ability to tag and overexpress proteins will greatly simplify purification of proteins in their native form (1, 3, 13). In addition, targeted gene deletion will allow the creation of mutants that can reveal in-vivo function of the candidate genes. The complete genome sequence of this strain will aid all future efforts of genetic manipulation.

P. FURIOSUS COM1 GENOME SEQUENCE

Phylogenetic analyses of hyperthermophilic archaea reveal a position near the base of the tree of life. While archaea are increasingly found in environments which are not considered extreme,

current evidence suggests that the ancestor of all extant Archaea was hyperthermophilic (9). These organisms are sometimes portrayed as relict species thriving in limited niches similar to the early earth, undergoing little change in the last few billion years. While this may have an element of truth, extensive genetic exchange and rearrangement in hydrothermal vent environments has been previously described (11, 19). Comparative genomic analyses have also revealed exchange of protein superfamilies from the archaea to prokaryotes and eukaryotes and vice versa (6). The sequencing of the COM1 strain of *P. furiosus* demonstrated that significant genetic change and large-scale rearrangement can occur over a short time in a laboratory environment (2). Examination of the genome arrangement of the publicly available type strain also showed rearrangement relative to the published reference sequence that was based on a strain maintained in a fashion similar to COM1 for several years before the sequencing project. The chromosomal sequence in NCBI therefore represents a variant laboratory strain which is not available for further research. Recent studies have investigated the evolution of strains in laboratory environments revealing significant variation (14, 18). The ever decreasing cost of genome sequencing will allow more extensive genotyping of both laboratory strains and environmental isolates leading to improved estimates of natural diversity and rates of change. Of particular interest will be the further investigation of other Thermococcaceae to determine whether the properties observed in COM1 are widespread in this family or unique to *Pyrococcus*.

METAL UPTAKE IN *D. VULGARIS* AND *P. FERMENTANS*

The metallomic study of *D. vulgaris* and *P. fermentans* found fundamental differences in metal uptake in these organisms. Further research will be required to elucidate the mechanisms by which metals are taken up, which proteins bind particular metals, and the mechanisms of

detoxification. Fractionation of cytoplasmic extracts of these two organisms via column chromatography is planned to estimate the number of novel metalloproteins which may be present, with the ultimate goal of purification and characterization. In particular, the metal-based purification of observed uranium peaks in *P. fermentans* is an important target for determining possible mechanisms of sequestration. The release of finished genome sequences for several *P. fermentans* strains will have a number of advantages in the study of this species. Microarray construction is planned, which will enable transcriptomic studies under relevant growth conditions, such as supplementation with toxic metals. The next generation GeoChip will also incorporate additional probes representing *P. fermentans* genes allowing the detection of genes and pathways important to this organisms in environmental samples (12). Additionally, the availability of complete genomes of strains that reduce uranium and those that do not will enable comparative genomic analysis to identify pathways involved in this process.

REFERENCES

1. **Basen, M., J. Sun, and M. W. Adams.** 2012. Engineering a hyperthermophilic archaeon for temperature-dependent product formation. MBio **3:**e00053-12.

2. **Bridger, S. L., W. A. Lancaster, F. L. Poole, 2nd, G. J. Schut, and M. W. Adams.** 2012. Genome Sequencing of a Genetically-Tractable *Pyrococcus furiosus* Strain Reveals a Highly Dynamic Genome. J Bacteriol.

3. **Chandrayan, S. K., P. M. McTernan, R. C. Hopkins, J. Sun, F. E. Jenney, Jr., and M. W. Adams.** 2012. Engineering Hyperthermophilic Archaeon *Pyrococcus furiosus* to Overproduce Its Cytoplasmic [NiFe]-Hydrogenase. J Biol Chem **287:**3257-64.

4. **Cvetkovic, A., A. L. Menon, M. P. Thorgersen, J. W. Scott, F. L. Poole, II, 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. W. Adams.** 2010. Microbial metalloproteomes are largely uncharacterized. Nature (London, U. K.) **466:**779-782.

5. **Dana, G. V., T. Kuiken, D. Rejeski, and A. A. Snow.** 2012. Synthetic biology: Four steps to avoid a synthetic-biology disaster. Nature **483:**29.

6. **David, L. A., and E. J. Alm.** 2011. Rapid evolutionary innovation during an Archaean genetic expansion. Nature **469:**93-6.

7. **Erickson, B., R. Singh, and P. Winters.** 2011. Synthetic biology: regulating industry uses of new biotechnologies. Science **333:**1254-6.

8. **Gibson, D. G., J. I. Glass, C. Lartigue, V. N. Noskov, R. Y. Chuang, M. A. Algire, G. A. Benders, M. G. Montague, L. Ma, M. M. Moodie, C. Merryman, S. Vashee, R. Krishnakumar, N. Assad-Garcia, C. Andrews-Pfannkoch, E. A. Denisova, L. Young, Z. Q. Qi, T. H. Segall-Shapiro, C. H. Calvey, P. P. Parmar, C. A. Hutchison, 3rd, H. O. Smith, and J. C. Venter.** 2010. Creation of a bacterial cell controlled by a chemically synthesized genome. Science **329:**52-6.

9. **Gribaldo, S., and C. Brochier-Armanet.** 2006. The origin and evolution of Archaea: a state of the art. Philos Trans R Soc Lond B Biol Sci **361:**1007-22.

10. **Gutmann, A.** 2011. The ethics of synthetic biology: guiding principles for emerging technologies. Hastings Cent Rep **41:**17-22.

11. **Hamilton-Brehm, S. D., G. J. Schut, and M. W. Adams.** 2005. Metabolic and evolutionary relationships among *Pyrococcus* Species: genetic exchange within a hydrothermal vent environment. J Bacteriol **187:**7492-9.

12. **He, Z., Y. Deng, J. D. Van Nostrand, Q. Tu, M. Xu, C. L. Hemme, X. Li, L. Wu, T. J. Gentry, Y. Yin, J. Liebich, T. C. Hazen, and J. Zhou.** 2010. GeoChip 3.0 as a high-throughput tool for analyzing microbial community composition, structure and functional activity. ISME J **4:**1167-79.

13. **Hopkins, R. C., J. Sun, F. E. Jenney, Jr., S. K. Chandrayan, P. M. McTernan, and M. W. Adams.** 2011. Homologous expression of a subcomplex of *Pyrococcus furiosus* hydrogenase that interacts with pyruvate ferredoxin oxidoreductase. PLoS One **6:**e26569, Oct 24 (epub).

14. **Ioerger, T. R., Y. Feng, K. Ganesula, X. Chen, K. M. Dobos, S. Fortune, W. R. Jacobs, Jr., V. Mizrahi, T. Parish, E. Rubin, C. Sassetti, and J. C. Sacchettini.** 2010. Variation among genome sequences of H37Rv strains of Mycobacterium tuberculosis from multiple laboratories. J Bacteriol **192:**3645-53.

15. **Lancaster, W. A., J. L. Praissman, F. L. Poole, 2nd, A. Cvetkovic, A. L. Menon, J. W. Scott, F. E. Jenney, Jr., M. P. Thorgersen, E. Kalisiak, J. V. Apon, S. A. Trauger, G. Siuzdak, J. A. Tainer, and M. W. Adams.** 2011. A computational framework for proteome-wide pursuit and prediction of metalloproteins using ICP-MS and MS/MS data. BMC Bioinformatics **12:**64.

16. **Lipscomb, G. L., K. Stirrett, G. J. Schut, F. Yang, F. E. Jenney, Jr., R. A. Scott, M. W. Adams, and J. Westpheling.** 2011. Natural competence in the hyperthermophilic archaeon *Pyrococcus furiosus* facilitates genetic manipulation: construction of markerless deletions of genes encoding the two cytoplasmic hydrogenases. Appl Environ Microbiol **77:**2232-8.

17. **Luisi, P. L., and P. Stano.** 2011. Synthetic biology: minimal cell mimicry. Nat Chem **3:**755-6.

18. **Srivatsan, A., Y. Han, J. Peng, A. K. Tehranchi, R. Gibbs, J. D. Wang, and R. Chen.** 2008. High-precision, whole-genome sequencing of laboratory strains facilitates genetic studies. PLoS Genet **4:**e1000139.

19. **White, J. R., P. Escobar-Paramo, E. F. Mongodin, K. E. Nelson, and J. DiRuggiero.** 2008. Extensive genome rearrangements and multiple horizontal gene transfers in a

population of *Pyrococcus* isolates from Vulcano Island, Italy. Appl Environ Microbiol **74:**6447-51.

REFERENCES

1. **Aiuppa, A., G. Dongarra, G. Capasso, and P. Allard.** 2000. Trace elements in the thermal groundwaters of Vulcano Island (Sicily). Journal of Volcanology and Geothermal Research **98:**189-207.

2. **Andreini, C., I. Bertini, G. Cavallaro, G. L. Holliday, and J. M. Thornton.** 2008. Metal ions in biological catalysis: from enzyme databases to general principles. Journal Of Biological Inorganic Chemistry **13:**1205-1218.

3. **Andreini, C., I. Bertini, and A. Rosato.** 2009. Metalloproteomes: A Bioinformatic Approach. Acc Chem Res **42:**1471-1479.

4. **Babor, M., S. Gerzon, B. Raveh, V. Sobolev, and M. Edelman.** 2008. Prediction of transition metal-binding sites from apo protein structures. Proteins **70:**208-17.

5. **Basen, M., J. Sun, and M. W. Adams.** 2012. Engineering a hyperthermophilic archaeon for temperature-dependent product formation. MBio **3:**e00053-12.

6. **Bertini, I., and G. Cavallaro.** 2010. Bioinformatics in bioinorganic chemistry. Metallomics **2:**39-51.

7. **Biswas, K. C., N. A. Woodards, H. Xu, and L. L. Barton.** 2009. Reduction of molybdate by sulfate-reducing bacteria. Biometals **22:**131-9.

8. **Blanchard, A., and C. Bebear.** 2012. The evolution of *Mycoplasma genitalium*. Ann N Y Acad Sci **1230:**E61-4.

9. **Blattner, F. R., G. Plunkett, III, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao.** 1997. The Complete Genome Sequence of Escherichia coli K-12. Science **277:**1453-1462.

10. **Boetzer, M., C. V. Henkel, H. J. Jansen, D. Butler, and W. Pirovano.** 2010. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics **27:**578-9.

11. **Brandis, A., and R. K. Thauer.** 1981. Growth of *Desulfovibrio* species on Hydrogen and Sulphate as Sole Energy Source. Journal of General Microbiology **126:**249-252.

12. **Bridger, S. L., S. M. Clarkson, K. Stirrett, M. B. DeBarry, G. L. Lipscomb, G. J. Schut, J. Westpheling, R. A. Scott, and M. W. Adams.** 2011. Deletion strains reveal metabolic roles for key elemental sulfur-responsive proteins in *Pyrococcus furiosus*. J Bacteriol **193:**6498-504.

13. **Bridger, S. L., W. A. Lancaster, F. L. Poole, 2nd, G. J. Schut, and M. W. Adams.** 2012. Genome Sequencing of a Genetically-Tractable *Pyrococcus furiosus* Strain Reveals a Highly Dynamic Genome. J Bacteriol.

14. **Brondino, C. D., M. J. Romao, I. Moura, and J. J. Moura.** 2006. Molybdenum and tungsten enzymes: the xanthine oxidase family. Curr Opin Chem Biol **10:**109-14.

15. **Bult, C. J., O. White, G. J. Olsen, L. Zhou, R. D. Fleischmann, G. G. Sutton, J. A. Blake, L. M. FitzGerald, R. A. Clayton, J. D. Gocayne, A. R. Kerlavage, B. A. Dougherty, J.-F. Tomb, M. D. Adams, C. I. Reich, R. Overbeek, E. F. Kirkness, K. G. Weinstock, J. M. Merrick, A. Glodek, J. L. Scott, N. S. M. Geoghagen, J. F. Weidman, J. L. Fuhrmann, D. Nguyen, T. R. Utterback, J. M. Kelley, J. D. Peterson, P. W. Sadow, M. C. Hanna, M. D. Cotton, K. M. Roberts, M. A. Hurst, B. P. Kaine, M. Borodovsky, H.-P. Klenk, C. M. Fraser, H. O. Smith, C. R. Woese, and J. C. Venter.** 1996. Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus jannaschii. Science **273:**1058-1073.

16. **Cabrera, G., R. Perez, J. M. Gomez, A. Abalos, and D. Cantero.** 2006. Toxic effects of dissolved heavy metals on *Desulfovibrio vulgaris* and *Desulfovibrio* sp. strains. J Hazard Mater **135:**40-6.

17. **Caffrey, S. M., and G. Voordouw.** 2010. Effect of sulfide on growth physiology and gene expression of *Desulfovibrio vulgaris* Hildenborough. Antonie Van Leeuwenhoek **97:**11-20.

18. **Carbone, A.** 2006. Computational prediction of genomic functional cores specific to different microbes. J Mol Evol **63:**733-46.

19. **Chandrayan, S. K., P. M. McTernan, R. C. Hopkins, J. Sun, F. E. Jenney, Jr., and M. W. Adams.** 2012. Engineering Hyperthermophilic Archaeon *Pyrococcus furiosus* to Overproduce Its Cytoplasmic [NiFe]-Hydrogenase. J Biol Chem **287:**3257-64.

20. **Chardin, B., A. Dolla, F. Chaspoul, M. L. Fardeau, P. Gallice, and M. Bruschi.** 2002. Bioremediation of chromate: thermodynamic analysis of the effects of Cr(VI) on sulfate-reducing bacteria. Appl Microbiol Biotechnol **60:**352-60.

21. **Chardin, B., M. T. Giudici-Orticoni, G. De Luca, B. Guigliarelli, and M. Bruschi.** 2003. Hydrogenases in sulfate-reducing bacteria function as chromium reductase. Appl Microbiol Biotechnol **63:**315-21.

22. **Chevreux, B.** 2005. MIRA: An Automated Genome and EST Assembler. German Cancer Research Center Heidelberg.

23. **Chhabra, S. R., G. Butland, D. A. Elias, J. M. Chandonia, O. Y. Fok, T. R. Juba, A. Gorur, S. Allen, C. M. Leung, K. L. Keller, S. Reveco, G. M. Zane, E. Semkiw, R. Prathapam, B. Gold, M. Singer, M. Ouellet, E. D. Szakal, D. Jorgens, M. N. Price, H. E. Witkowska, H. R. Beller, A. P. Arkin, T. C. Hazen, M. D. Biggin, M. Auer, J. D. Wall, and J. D. Keasling.** 2011. Generalized schemes for high-throughput manipulation of the *Desulfovibrio vulgaris* genome. Appl Environ Microbiol **77:**7595-604.

24. **Chinen, A., I. Uchiyama, and I. Kobayashi.** 2000. Comparison between *Pyrococcus horikoshii* and *Pyrococcus abyssi* genome sequences reveals linkage of restriction-modification genes with large genome polymorphisms. Gene **259:**109-21.

25. **Chvatal, V.** 1979. A Greedy Heuristic for the Set-Covering Problem. MATHEMATICS OF OPERATIONS RESEARCH **4:**233-235.

26. **Clark, M. E., Q. He, Z. He, K. H. Huang, E. J. Alm, X. F. Wan, T. C. Hazen, A. P. Arkin, J. D. Wall, J. Z. Zhou, and M. W. Fields.** 2006. Temporal transcriptomic analysis as Desulfovibrio vulgaris Hildenborough transitions into stationary phase during electron donor depletion. Appl Environ Microbiol **72:**5578-88.

27. **Clark, W. T., and P. Radivojac.** 2011. Analysis of protein function and its prediction from amino acid sequence. Proteins **79:**2086-96.

28. **Clarkson, S. M.** 2011. Effects of Sulfide on Iron Metabolism of the Hyperthermophilic Archaeon *Pyrococcus furiosus*. University of Georgia, Athens.

29. **Cvetkovic, A., A. L. Menon, M. P. Thorgersen, J. W. Scott, F. L. Poole, II, 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. W. Adams.** 2010. Microbial metalloproteomes are largely uncharacterized. Nature (London, U. K.) **466:**779-782.

30. **da Silva, S. M., C. Pimentel, F. M. Valente, C. Rodrigues-Pousada, and I. A. Pereira.** 2011. Tungsten and molybdenum regulation of formate dehydrogenase expression in *Desulfovibrio vulgaris* Hildenborough. J Bacteriol **193:**2909-16.

31. **Dana, G. V., T. Kuiken, D. Rejeski, and A. A. Snow.** 2012. Synthetic biology: Four steps to avoid a synthetic-biology disaster. Nature **483:**29.

32. **David, L. A., and E. J. Alm.** 2011. Rapid evolutionary innovation during an Archaean genetic expansion. Nature **469:**93-6.

33. **Deckert, G., P. V. Warren, T. Gaasterland, W. G. Young, A. L. Lenox, D. E. Graham, R. Overbeek, M. A. Snead, M. Keller, M. Aujay, R. Huber, R. A. Feldman, J. M. Short, G. J. Olsen, and R. V. Swanson.** 1998. The complete genome of the hyperthermophilic bacterium Aquifex aeolicus. Nature **392:**353-8.

34. **Deland, M. R.** 1981. Regulatory alert: superfund. Environ Sci Technol **15:**255.

35. **Dewall, M. T., and D. W. Cheng.** 2011. The minimal genome: a metabolic and environmental comparison. Brief Funct Genomics **10:**312-5.

36. **Dudev, T., and C. Lim.** 2008. Metal Binding Affinity and Selectivity in Metalloproteins: Insights from Computational Studies. Annual Review of Biophysics **37:**97-116.

37. **Dupont, C. L., A. Butcher, R. E. Valas, P. E. Bourne, and G. Caetano-Anolles.** 2010. History of biological metal utilization inferred through phylogenomic analysis of protein structures. Proc Natl Acad Sci U S A **107:**10567-72.

38. **Edgcomb, V. P., S. J. Molyneaux, M. A. Saito, K. Lloyd, S. Boer, C. O. Wirsen, M. S. Atkins, and A. Teske.** 2004. Sulfide ameliorates metal toxicity for deep-sea hydrothermal vent archaea. Appl Environ Microbiol **70:**2551-5.

39. **Eguchi, M., T. Nishikawa, K. Macdonald, R. Cavicchioli, J. C. Gottschal, and S. Kjelleberg.** 1996. Responses to Stress and Nutrient Availability by the Marine Ultramicrobacterium Sphingomonas sp. Strain RB2256. Appl Environ Microbiol **62:**1287-1294.

40. **Eidsness, M. K., S. E. O'Dell, D. M. Kurtz, Jr., R. L. Robson, and R. A. Scott.** 1992. Expression of a synthetic gene coding for the amino acid sequence of *Clostridium pasteurianum* rubredoxin. Protein Eng **5:**367-71.

41. **Elias, D. A., J. M. Suflita, M. J. McInerney, and L. R. Krumholz.** 2004. Periplasmic cytochrome c3 of *Desulfovibrio vulgaris* is directly involved in H2-mediated metal but not sulfate reduction. Appl Environ Microbiol **70:**413-20.

42. **Elowitz, M., and W. A. Lim.** 2010. Build life to understand it. Nature **468:**889-90.

43. **Erickson, B., R. Singh, and P. Winters.** 2011. Synthetic biology: regulating industry uses of new biotechnologies. Science **333:**1254-6.

44. **Farkas, J., D. Chung, M. DeBarry, M. W. Adams, and J. Westpheling.** 2011. Defining components of the chromosomal origin of replication of the hyperthermophilic archaeon

Pyrococcus furiosus needed for construction of a stable replicating shuttle vector. Appl Environ Microbiol **77:**6343-9.

45. **Farkas, J., K. Stirrett, G. L. Lipscomb, W. Nixon, R. A. Scott, M. W. Adams, and J. Westpheling.** 2012. Recombinogenic Properties of *Pyrococcus furiosus* Strain COM1 Enable Rapid Selection of Targeted Mutants. Appl Environ Microbiol **78:**4669-76.

46. **Fiala, G., and K. Stetter.** 1986. *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100°C. Arch Microbiol **145:**56-61.

47. **Figueiredo, M. C., S. A. Lobo, J. N. Carita, L. S. Nobre, and L. M. Saraiva.** 2012. Bacterioferritin protects the anaerobe *Desulfovibrio vulgaris* Hildenborough against oxygen. Anaerobe.

48. **Fisunov, G. Y., D. G. Alexeev, N. A. Bazaleev, V. G. Ladygina, M. A. Galyamina, I. G. Kondratov, N. A. Zhukova, M. V. Serebryakova, I. A. Demina, and V. M. Govorun.** 2011. Core proteome of the minimal cell: comparative proteomics of three mollicute species. PLoS One **6:**e21964.

49. **Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J. F. Tomb, B. A. Dougherty, J. M. Merrick, and et al.** 1995. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science **269:**496-512.

50. **Forterre, P., S. Gribaldo, and C. Brochier-Armanet.** 2009. Happy together: genomic insights into the unique Nanoarchaeum/Ignicoccus association. J Biol **8:**7.

51. **Fraser, C. M., J. D. Gocayne, O. White, M. D. Adams, R. A. Clayton, R. D. Fleischmann, C. J. Bult, A. R. Kerlavage, G. Sutton, J. M. Kelley, R. D. Fritchman, J. F. Weidman, K. V. Small, M. Sandusky, J. Fuhrmann, D. Nguyen, T. R. Utterback, D. M. Saudek, C. A. Phillips, J. M. Merrick, J. F. Tomb, B. A. Dougherty, K. F. Bott, P. C. Hu, T. S. Lucier, S. N. Peterson, H. O. Smith, C. A. Hutchison, 3rd, and J. C. Venter.** 1995. The minimal gene complement of Mycoplasma genitalium. Science **270:**397-403.

52. **Garcia, J. S., C. S. Magalhaes, and M. A. Arruda.** 2006. Trends in metal-binding and metalloprotein analysis. Talanta **69:**1-15.

53. **Gibson, D. G., G. A. Benders, C. Andrews-Pfannkoch, E. A. Denisova, H. Baden-Tillson, J. Zaveri, T. B. Stockwell, A. Brownley, D. W. Thomas, M. A. Algire, C. Merryman, L. Young, V. N. Noskov, J. I. Glass, J. C. Venter, C. A. Hutchison, 3rd, and H. O. Smith.** 2008. Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. Science **319:**1215-20.

54. **Gibson, D. G., J. I. Glass, C. Lartigue, V. N. Noskov, R. Y. Chuang, M. A. Algire, G. A. Benders, M. G. Montague, L. Ma, M. M. Moodie, C. Merryman, S. Vashee, R. Krishnakumar, N. Assad-Garcia, C. Andrews-Pfannkoch, E. A. Denisova, L. Young, Z. Q. Qi, T. H. Segall-Shapiro, C. H. Calvey, P. P. Parmar, C. A. Hutchison, 3rd, H. O. Smith, and J. C. Venter.** 2010. Creation of a bacterial cell controlled by a chemically synthesized genome. Science **329:**52-6.

55. **Gihring, T. M., G. Zhang, C. C. Brandt, S. C. Brooks, J. H. Campbell, S. Carroll, C. S. Criddle, S. J. Green, P. Jardine, J. E. Kostka, K. Lowe, T. L. Mehlhorn, W. Overholt, D. B. Watson, Z. Yang, W. M. Wu, and C. W. Schadt.** 2011. A limited microbial consortium is responsible for extended bioreduction of uranium in a contaminated aquifer. Appl Environ Microbiol **77:**5955-65.

56. **Gilmour, C. C., D. A. Elias, A. M. Kucken, S. D. Brown, A. V. Palumbo, C. W. Schadt, and J. D. Wall.** 2011. The Sulfate-reducing bacterium *Desulfovibrio desulfuricans* ND132 as a model for understanding bacterial mercury methylation. Appl Environ Microbiol **77:**3938-51.

57. **Gladyshev, V. N., S. V. Khangulov, M. J. Axley, and T. C. Stadtman.** 1994. Coordination of selenium to molybdenum in formate dehydrogenase H from Escherichia coli. Proc Natl Acad Sci U S A **91:**7708-11.

58. **Glass, J. I., N. Assad-Garcia, N. Alperovich, S. Yooseph, M. R. Lewis, M. Maruf, C. A. Hutchison, 3rd, H. O. Smith, and J. C. Venter.** 2006. Essential genes of a minimal bacterium. Proc Natl Acad Sci U S A **103:**425-30.

59. **Golden, D. C., D. W. Ming, R. V. Morris, A. J. Brearley, H. V. Lauer, Jr., A. H. Treiman, M. E. Zolensky, C. S. Schwandt, G. E. Lofgren, and G. A. McKay.** 2004. Evidence for exclusively inorganic formation of magnetite in Martian meteorite ALH84001. American Mineralogist **89:**681-695.

60. **Gonzaga-Jauregui, C., J. R. Lupski, and R. A. Gibbs.** 2012. Human genome sequencing in health and disease. Annu Rev Med **63:**35-61.

61. **Goulhen, F., A. Gloter, F. Guyot, and M. Bruschi.** 2006. Cr(VI) detoxification by *Desulfovibrio vulgaris* strain Hildenborough: microbe-metal interactions studies. Appl Microbiol Biotechnol **71:**892-7.

62. **Gribaldo, S., and C. Brochier-Armanet.** 2006. The origin and evolution of Archaea: a state of the art. Philos Trans R Soc Lond B Biol Sci **361:**1007-22.

63. **Gutmann, A.** 2011. The ethics of synthetic biology: guiding principles for emerging technologies. Hastings Cent Rep **41:**17-22.

64. **Haferburg, G., and E. Kothe.** 2010. Metallomics: lessons for metalliferous soil remediation. Appl Microbiol Biotechnol **87:**1271-80.

65. **Hamilton-Brehm, S. D., G. J. Schut, and M. W. Adams.** 2005. Metabolic and evolutionary relationships among *Pyrococcus* Species: genetic exchange within a hydrothermal vent environment. J Bacteriol **187:**7492-9.

66. **Hansel, C. M., S. Fendorf, P. M. Jardine, and C. A. Francis.** 2008. Changes in bacterial and archaeal community structure and functional diversity along a geochemically variable soil profile. Appl Environ Microbiol **74:**1620-33.

67. **Haschke, R. H., and L. L. Campbell.** 1971. Thiosulfate reductase of *Desulfovibrio vulgaris*. J Bacteriol **106:**603-7.

68. **He, Z., Y. Deng, J. D. Van Nostrand, Q. Tu, M. Xu, C. L. Hemme, X. Li, L. Wu, T. J. Gentry, Y. Yin, J. Liebich, T. C. Hazen, and J. Zhou.** 2010. GeoChip 3.0 as a high-throughput tool for analyzing microbial community composition, structure and functional activity. ISME J **4:**1167-79.

69. **He, Z., A. Zhou, E. Baidoo, Q. He, M. P. Joachimiak, P. Benke, R. Phan, A. Mukhopadhyay, C. L. Hemme, K. Huang, E. J. Alm, M. W. Fields, J. Wall, D. Stahl, T. C. Hazen, J. D. Keasling, A. P. Arkin, and J. Zhou.** 2009. Global transcriptional, physiological, and metabolite analyses of the responses of *Desulfovibrio vulgaris* hildenborough to salt adaptation. Appl Environ Microbiol **76:**1574-86.

70. **Heidelberg, J. F., R. Seshadri, S. A. Haveman, C. L. Hemme, I. T. Paulsen, J. F. Kolonay, J. A. Eisen, N. Ward, B. Methe, L. M. Brinkac, S. C. Daugherty, R. T. Deboy, R. J. Dodson, A. S. Durkin, R. Madupu, W. C. Nelson, S. A. Sullivan, D. Fouts, D. H. Haft, J. Selengut, J. D. Peterson, T. M. Davidsen, N. Zafar, L. Zhou, D. Radune, G. Dimitrov, M. Hance, K. Tran, H. Khouri, J. Gill, T. R. Utterback, T. V. Feldblyum, J. D. Wall, G. Voordouw, and C. M. Fraser.** 2004. The genome sequence of the anaerobic, sulfate-reducing bacterium *Desulfovibrio vulgaris* Hildenborough. Nat Biotechnol **22:**554-9.

71. **Holden, J. F., F. L. Poole Ii, S. L. Tollaksen, C. S. Giometti, H. Lim, J. R. Yates Iii, and M. W. Adams.** 2001. Identification of membrane proteins in the hyperthermophilic archaeon *Pyrococcus furiosus* using proteomics and prediction programs. Comp Funct Genomics **2:**275-88.

72. **Hopkins, R. C., J. Sun, F. E. Jenney, Jr., S. K. Chandrayan, P. M. McTernan, and M. W. Adams.** 2011. Homologous expression of a subcomplex of *Pyrococcus furiosus* hydrogenase that interacts with pyruvate ferredoxin oxidoreductase. PLoS One **6:**e26569, Oct 24 (epub).

73. **Houston, M. C.** 2011. Role of mercury toxicity in hypertension, cardiovascular disease, and stroke. J Clin Hypertens (Greenwich) **13:**621-7.

74. **Huber, H., M. J. Hohn, R. Rachel, T. Fuchs, V. C. Wimmer, and K. O. Stetter.** 2002. A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. Nature **417:**63-7.

75. **Humphries, A. C., K. P. Nott, L. D. Hall, and L. E. Macaskie.** 2004. Continuous removal of Cr(VI) from aqueous solution catalysed by palladised biomass of *Desulfovibrio vulgaris*. Biotechnol Lett **26:**1529-32.

76. **Hunter, S., R. Apweiler, T. K. Attwood, A. Bairoch, A. Bateman, D. Binns, P. Bork, U. Das, L. Daugherty, L. Duquenne, R. D. Finn, J. Gough, D. Haft, N. Hulo, D. Kahn, E. Kelly, A. Laugraud, I. Letunic, D. Lonsdale, R. Lopez, M. Madera, J. Maslen, C. McAnulla, J. McDowall, J. Mistry, A. Mitchell, N. Mulder, D. Natale, C. Orengo, A. F. Quinn, J. D. Selengut, C. J. Sigrist, M. Thimma, P. D. Thomas, F. Valentin, D. Wilson, C. H. Wu, and C. Yeats.** 2009. InterPro: the integrative protein signature database. Nucleic Acids Res **37:**D211-5.

77. **Ioerger, T. R., Y. Feng, K. Ganesula, X. Chen, K. M. Dobos, S. Fortune, W. R. Jacobs, Jr., V. Mizrahi, T. Parish, E. Rubin, C. Sassetti, and J. C. Sacchettini.** 2010. Variation among genome sequences of H37Rv strains of Mycobacterium tuberculosis from multiple laboratories. J Bacteriol **192:**3645-53.

78. **Jarup, L.** 2003. Hazards of heavy metal contamination. Br Med Bull **68:**167-82.

79. **Jenney, F. E., Jr., and M. W. Adams.** 2001. Rubredoxin from Pyrococcus furiosus. Methods Enzymol **334:**45-55.

80. **Jomova, K., and M. Valko.** 2011. Advances in metal-induced oxidative stress and human disease. Toxicology **283:**65-87.

81. **Jun, X., L. Lupeng, X. Minjuan, P. Oger, W. Fengping, M. Jebbar, and X. Xiang.** 2011. Complete genome sequence of the obligate piezophilic hyperthermophilic archaeon *Pyrococcus yayanosii* CH1. J Bacteriol **193:**4297-8.

82. **Kajander, E. O., I. Kuronen, K. K. Akerman, A. Pelttari, and N. Ciftcioglu.** 1997. Presented at the Instruments, Methods, and Missions for the Investigation of Extraterrestrial Microorganisms, San Diego, CA, USA.

83. **Kanoksilapatham, W., J. M. Gonzalez, D. L. Maeder, J. DiRuggiero, and F. T. Robb.** 2004. A proposal to rename the hyperthermophile *Pyrococcus woesei* as *Pyrococcus furiosus* subsp. *woesei*. Archaea **1:**277-83.

84. **Kawarabayasi, Y., M. Sawada, H. Horikawa, Y. Haikawa, Y. Hino, S. Yamamoto, M. Sekine, S. Baba, H. Kosugi, A. Hosoyama, Y. Nagai, M. Sakai, K. Ogura, R. Otsuka, H. Nakazawa, M. Takamiya, Y. Ohfuku, T. Funahashi, T. Tanaka, Y. Kudoh, J. Yamazaki, N.**

Kushida, A. Oguchi, K. Aoki, and H. Kikuchi. 1998. Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium, *Pyrococcus horikoshii* OT3. DNA Res **5:**55-76.

85. **Kehl, C., A. M. Simms, R. D. Toofanny, and V. Daggett.** 2008. Dynameomics: a multi-dimensional analysis-optimized database for dynamic protein data. Protein Eng Des Sel **21:**379-86.

86. **Kengen, S. W., E. J. Luesink, A. J. Stams, and A. J. Zehnder.** 1993. Purification and characterization of an extremely thermostable beta-glucosidase from the hyperthermophilic archaeon *Pyrococcus furiosus*. Eur J Biochem **213:**305-12.

87. **Kim, H. W., and K. Ishikawa.** 2010. Complete saccharification of cellulose at high temperature using endocellulase and beta-glucosidase from *Pyrococcus* sp. J Microbiol Biotechnol **20:**889-92.

88. **Kletzin, A., and M. W. Adams.** 1996. Tungsten in biological systems. FEMS Microbiol Rev **18:**5-63.

89. **Klonowska, A., M. E. Clark, S. B. Thieman, B. J. Giles, J. D. Wall, and M. W. Fields.** 2008. Hexavalent chromium reduction in *Desulfovibrio vulgaris* Hildenborough causes transitory inhibition of sulfate reduction and cell growth. Appl Microbiol Biotechnol **78:**1007-16.

90. **Lamrabet, O., L. Pieulle, C. Aubert, F. Mouhamar, P. Stocker, A. Dolla, and G. Brasseur.** 2011. Oxygen reduction in the strict anaerobe *Desulfovibrio vulgaris* Hildenborough: characterization of two membrane-bound oxygen reductases. Microbiology **157:**2720-32.

91. **Lancaster, W. A., and M. W. W. Adams.** 2011. The Influence of Environment and Metabolic Capacity on the Size of a Microorganism, p. 93-103. *In* P. L. Luisi and P. Stano (ed.), The Minimal Cell: The Biophysics of Cell Compartment and the Origin of Cell Functionality. Springer Netherlands.

92. **Lancaster, W. A., J. L. Praissman, F. L. Poole, 2nd, A. Cvetkovic, A. L. Menon, J. W. Scott, F. E. Jenney, Jr., M. P. Thorgersen, E. Kalisiak, J. V. Apon, S. A. Trauger, G. Siuzdak, J. A. Tainer, and M. W. Adams.** 2011. A computational framework for proteome-wide pursuit and prediction of metalloproteins using ICP-MS and MS/MS data. BMC Bioinformatics **12:**64.

93. **Langfelder, P., B. Zhang, and S. Horvath.** 2008. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics **24:**719-20.

94. **Lapierre, P., and J. P. Gogarten.** 2009. Estimating the size of the bacterial pan-genome. Trends Genet **25:**107-10.

95. **Lecompte, O., R. Ripp, V. Puzos-Barbe, S. Duprat, R. Heilig, J. Dietrich, J. C. Thierry, and O. Poch.** 2001. Genome evolution at the genus level: comparison of three complete genomes of hyperthermophilic archaea. Genome Res **11:**981-93.

96. **Lee, A. M., J. R. Sevinsky, J. L. Bundy, A. M. Grunden, and J. L. Stephenson, Jr.** 2009. Proteomics of *Pyrococcus furiosus*, a hyperthermophilic archaeon refractory to traditional methods. J Proteome Res **8:**3844-51.

97. **Lin, X., J. McKinley, C. T. Resch, R. Kaluzny, C. L. Lauber, J. Fredrickson, R. Knight, and A. Konopka.** 2012. Spatial and temporal dynamics of the microbial community in the Hanford unconfined aquifer. ISME J.

98. **Liolios, K., K. Mavromatis, N. Tavernarakis, and N. C. Kyrpides.** 2008. The Genomes On Line Database (GOLD) in 2007: status of genomic and metagenomic projects and their associated metadata. Nucleic Acids Res **36:**D475-9.

99. **Lipscomb, G. L., K. Stirrett, G. J. Schut, F. Yang, F. E. Jenney, Jr., R. A. Scott, M. W. Adams, and J. Westpheling.** 2011. Natural competence in the hyperthermophilic archaeon *Pyrococcus furiosus* facilitates genetic manipulation: construction of markerless deletions of genes encoding the two cytoplasmic hydrogenases. Appl Environ Microbiol **77:**2232-8.

100. **Loveland-Curtze, J., V. I. Miteva, and J. E. Brenchley.** 2009. Herminiimonas glaciei sp. nov., a novel ultramicrobacterium from 3042 m deep Greenland glacial ice. Int J Syst Evol Microbiol **59:**1272-7.

101. **Lovley, D. R., and E. J. Phillips.** 1994. Reduction of Chromate by Desulfovibrio vulgaris and Its c(3) Cytochrome. Appl Environ Microbiol **60:**726-8.

102. **Lovley, D. R., P. K. Widman, J. C. Woodward, and E. J. Phillips.** 1993. Reduction of uranium by cytochrome c3 of *Desulfovibrio vulgaris*. Appl Environ Microbiol **59:**3572-6.

103. **Lu, P., C. Vogel, R. Wang, X. Yao, and E. M. Marcotte.** 2007. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. Nat Biotechnol **25:**117-24.

104. **Luisi, P. L., and P. Stano.** 2011. Synthetic biology: minimal cell mimicry. Nat Chem **3:**755-6.

105. **Manley, S. A., and J. Gailer.** 2009. Analysis of the plasma metalloproteome by SEC-ICP-AES: bridging proteomics and metabolomics. Expert Rev Proteomics **6:**251-65.

106. **Marchandin, H., C. Teyssier, J. Campos, H. Jean-Pierre, F. Roger, B. Gay, J. P. Carlier, and E. Jumas-Bilak.** 2010. Negativicoccus succinicivorans gen. nov., sp. nov., isolated from human clinical samples, emended description of the family *Veillonellaceae* and description of *Negativicutes* classis nov., *Selenomonadales* ord. nov. and *Acidaminococcaceae* fam. nov. in the bacterial phylum Firmicutes. Int J Syst Evol Microbiol **60:**1271-9.

107. **Marques, M. C., R. Coelho, A. L. De Lacey, I. A. Pereira, and P. M. Matias.** 2009. The three-dimensional structure of [NiFeSe] hydrogenase from *Desulfovibrio vulgaris* Hildenborough: a hydrogenase without a bridging ligand in the active site in its oxidised, "as-isolated" state. J Mol Biol **396:**893-907.

108. **Martel, J., and J. D. Young.** 2008. Purported nanobacteria in human blood as calcium carbonate nanoparticles. Proc Natl Acad Sci U S A **105:**5549-54.

109. **Martin, W., J. Baross, D. Kelley, and M. J. Russell.** 2008. Hydrothermal vents and the origin of life. Nat Rev Microbiol **6:**805-14.

110. **Martins, M., M. L. Faleiro, R. J. Barros, A. R. Verissimo, M. A. Barreiros, and M. C. Costa.** 2009. Characterization and activity studies of highly heavy metal resistant sulphate-reducing bacteria to be used in acid mine drainage decontamination. J Hazard Mater **166:**706-13.

111. **Martins, M., M. L. Faleiro, S. Chaves, R. Tenreiro, and M. C. Costa.** 2010. Effect of uranium (VI) on two sulphate-reducing bacteria cultures from a uranium mine site. Sci Total Environ **408:**2621-8.

112. **Martins, M., M. L. Faleiro, S. Chaves, R. Tenreiro, E. Santos, and M. C. Costa.** 2010. Anaerobic bio-removal of uranium (VI) and chromium (VI): comparison of microbial community structure. J Hazard Mater **176:**1065-72.

113. **McCliment, E. A., K. M. Voglesonger, P. A. O'Day, E. E. Dunn, J. R. Holloway, and S. C. Cary.** 2006. Colonization of nascent, deep-sea hydrothermal vents by a novel Archaeal and Nanoarchaeal assemblage. Environ Microbiol **8:**114-25.

114. **McKay, D. S., E. K. Gibson, Jr., K. L. Thomas-Keprta, H. Vali, C. S. Romanek, S. J. Clemett, X. D. Chillier, C. R. Maechling, and R. N. Zare.** 1996. Search for past life on Mars: possible relic biogenic activity in martian meteorite ALH84001. Science **273:**924-30.

115. **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.** 2009. Novel multiprotein complexes identified in the hyperthermophilic archaeon *Pyrococcus furiosus* by non-denaturing fractionation of the native proteome. Mol Cell Proteomics **8:**735-51.

116. **Moe, W. M., R. E. Stebbing, J. U. Rao, K. S. Bowman, M. F. Nobre, M. S. da Costa, and F. A. Rainey.** 2011. *Pelosinus defluvii* sp. nov., isolated from chlorinated solvent contaminated groundwater, emended description of the genus *Pelosinus*, and transfer of *Sporotalea propionica* to *Pelosinus propionicus* comb. nov. Int J Syst Evol Microbiol.

117. **Morris, J. J., R. E. Lenski, and E. R. Zinser.** 2012. The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss. MBio **3**.

118. **Mosher, J. J., T. J. Phelps, M. Podar, R. A. Hurt, Jr., J. H. Campbell, M. M. Drake, J. G. Moberly, C. W. Schadt, S. D. Brown, T. C. Hazen, A. P. Arkin, A. V. Palumbo, B. A. Faybishenko, and D. A. Elias.** 2012. Microbial community succession during lactate amendment and electron acceptor limitation reveals a predominance of metal-reducing *Pelosinus* spp. Appl Environ Microbiol **78:**2082-91.

119. **Mounicou, S., J. Szpunar, and R. Lobinski.** 2009. Metallomics: the concept and methodology. Chemical Society Reviews **38:**1119-1138.

120. **Mulder, N., and R. Apweiler.** 2007. InterPro and InterProScan: tools for protein sequence classification and comparison. Methods Mol Biol **396:**59-70.

121. **Mushegian, A. R., and E. V. Koonin.** 1996. A minimal gene set for cellular life derived by comparison of complete bacterial genomes. Proc Natl Acad Sci U S A **93:**10268-73.

122. **Nakabachi, A., A. Yamashita, H. Toh, H. Ishikawa, H. E. Dunbar, N. A. Moran, and M. Hattori.** 2006. The 160-kilobase genome of the bacterial endosymbiont Carsonella. Science **314:**267.

123. **Odom, J. M., and H. D. Peck, Jr.** 1984. Hydrogenase, electron-transfer proteins, and energy coupling in the sulfate-reducing bacteria *Desulfovibrio*. Annu Rev Microbiol **38:**551-92.

124. **Oslowski, D. M., J. H. Jung, D. H. Seo, C. S. Park, and J. F. Holden.** 2011. Production of hydrogen from alpha-1,4- and beta-1,4-linked saccharides by marine hyperthermophilic Archaea. Appl Environ Microbiol **77:**3169-73.

125. **Pankhania, I. P., A. N. Moosavi, and W. A. Hamilton.** 1986. Utilization of Cathodic Hydrogen by *Desulfovibrio vulgaris* (Hildenborough). Journal of General Microbiology **132:**3357-3365.

126. **Park, H. S., S. Lin, and G. Voordouw.** 2008. Ferric iron reduction by *Desulfovibrio vulgaris* Hildenborough wild type and energy metabolism mutants. Antonie Van Leeuwenhoek **93:**79-85.

127. **Passerini, A., M. Lippi, and P. Frasconi.** 2011. MetalDetector v2.0: predicting the geometry of metal binding sites from protein sequence. Nucleic Acids Res **39:**W288-92.

128. **Payne, R. B., D. M. Gentry, B. J. Rapp-Giles, L. Casalot, and J. D. Wall.** 2002. Uranium reduction by *Desulfovibrio desulfuricans* strain G20 and a cytochrome c3 mutant. Appl Environ Microbiol **68:**3129-32.

129. **Pereira, P. M., Q. He, F. M. Valente, A. V. Xavier, J. Zhou, I. A. Pereira, and R. O. Louro.** 2008. Energy metabolism in *Desulfovibrio vulgaris* Hildenborough: insights from transcriptome analysis. Antonie Van Leeuwenhoek **93:**347-62.

130. **Pietzsch, K., and W. Babel.** 2003. A sulfate-reducing bacterium that can detoxify U(VI) and obtain energy via nitrate reduction. J Basic Microbiol **43:**348-61.

131. **Piotrowska-Seget, Z., M. Cycoń, and J. Kozdrój.** 2005. Metal-tolerant bacteria occurring in heavily polluted soil and mine spoil. Applied Soil Ecology **28:**237-246.

132. **Poole, F. L., 2nd, B. A. Gerwe, R. C. Hopkins, G. J. Schut, M. V. Weinberg, F. E. Jenney, Jr., and M. W. Adams.** 2005. Defining genes in the genome of the hyperthermophilic archaeon *Pyrococcus furiosus*: implications for all microbial genomes. J Bacteriol **187:**7325-32.

133. **Porat, I., T. A. Vishnivetskaya, J. J. Mosher, C. C. Brandt, Z. K. Yang, S. C. Brooks, L. Liang, M. M. Drake, M. Podar, S. D. Brown, and A. V. Palumbo.** 2010. Characterization of archaeal community in contaminated and uncontaminated surface stream sediments. Microb Ecol **60:**784-95.

134. **Postgate, J. R.** 1951. On the nutrition of *Desulphovibrio desulphuricans*. J Gen Microbiol **5:**714-24.

135. **Postgate, J. R.** 1953. Presence of cytochrome in an obligate anaerobe. Biochem J **56:**xi-xii.

136. **Price, M. N., A. M. Deutschbauer, J. V. Kuehl, H. Liu, H. E. Witkowska, and A. P. Arkin.** 2010. Evidence-based annotation of transcripts and proteins in the sulfate-reducing bacterium *Desulfovibrio vulgaris* Hildenborough. J Bacteriol **193:**5716-27.

137. **Pruitt, K. D., T. Tatusova, and D. R. Maglott.** 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res **35:**D61-5.

138. **Punta, M., and Y. Ofran.** 2008. The rough guide to in silico function prediction, or how to use sequence and structure information to predict protein function. PLoS Comput Biol **4:**e1000160.

139. **Qafoku, N. P., P. Evan Dresel, E. Ilton, J. P. McKinley, and C. T. Resch.** 2010. Chromium transport in an acidic waste contaminated subsurface medium: the role of reduction. Chemosphere **81:**1492-500.

140. **R Development Core Team** 2005, posting date. R : A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. [Online.]

141. **Ramos, A. R., K. L. Keller, J. D. Wall, and I. A. Pereira.** 2012. The Membrane QmoABC Complex Interacts Directly with the Dissimilatory Adenosine 5'-Phosphosulfate Reductase in Sulfate Reducing Bacteria. Front Microbiol **3:**137.

142. **Raoult, D., and P. Forterre.** 2008. Redefining viruses: lessons from Mimivirus. Nat Rev Microbiol **6:**315-9.

143. **Ray, A. E., J. R. Bargar, V. Sivaswamy, A. C. Dohnalkova, Y. Fujita, B. M. Peyton, and T. S. Magnuson.** 2011. Evidence for multiple modes of uranium immobilization by an anaerobic bacterium. Geochimica et Cosmochimica Acta **75:**2684-2695.

144. **Ray, A. E., S. A. Connon, P. P. Sheridan, J. Gilbreath, M. Shields, D. T. Newby, Y. Fujita, and T. S. Magnuson.** 2010. Intragenomic heterogeneity of the 16S rRNA gene in strain UFO1 caused by a 100-bp insertion in helix 6. FEMS Microbiol Ecol **72:**343-53.

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

146. **Romao, M. J., J. Knablein, R. Huber, and J. J. Moura.** 1997. Structure and function of molybdopterin containing enzymes. Prog Biophys Mol Biol **68:**121-44.

147. **Schwarz, G., R. R. Mendel, and M. W. Ribbe.** 2009. Molybdenum cofactors, enzymes and pathways. Nature **460:**839-47.

148. **Sevcenco, A. M., G. C. Krijger, M. W. Pinkse, P. D. Verhaert, W. R. Hagen, and P. L. Hagedoorn.** 2009. Development of a generic approach to native metalloproteomics: application to the quantitative identification of soluble copper proteins in *Escherichia coli*. J Biol Inorg Chem **14:**631-40.

149. **Sevcenco, A. M., M. W. Pinkse, E. Bol, G. C. Krijger, H. T. Wolterbeek, P. D. Verhaert, P. L. Hagedoorn, and W. R. Hagen.** 2009. The tungsten metallome of *Pyrococcus furiosus*. Metallomics **1:**395-402.

150. **Shelobolina, E. S., K. P. Nevin, J. D. Blakeney-Hayward, C. V. Johnsen, T. W. Plaia, P. Krader, T. Woodard, D. E. Holmes, C. G. Vanpraagh, and D. R. Lovley.** 2007. *Geobacter pickeringii* sp. nov., *Geobacter argillaceus* sp. nov. and *Pelosinus fermentans* gen. nov., sp. nov., isolated from subsurface kaolin lenses. Int J Syst Evol Microbiol **57:**126-35.

151. **Shirai, M., T. Taniguchi, and H. Kambara.** 2012. Emerging Applications of Single-Cell Diagnostics. Top Curr Chem.

152. **Silva, J. C., M. V. Gorenstein, G. Z. Li, J. P. Vissers, and S. J. Geromanos.** 2006. Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. Mol Cell Proteomics **5:**144-56.

153. **Singh, S., S. H. Kang, A. Mulchandani, and W. Chen.** 2008. Bioremediation: environmental clean-up through pathway engineering. Curr Opin Biotechnol **19:**437-44.

154. **Sivaraman, J., R. S. Myers, L. Boju, T. Sulea, M. Cygler, V. Jo Davisson, and J. D. Schrag.** 2005. Crystal structure of Methanobacterium thermoautotrophicum phosphoribosyl-AMP cyclohydrolase HisI. Biochemistry **44:**10071-80.

155. **Sodhi, J. S., K. Bryson, L. J. McGuffin, J. J. Ward, L. Wernisch, and D. T. Jones.** 2004. Predicting metal-binding site residues in low-resolution structural models. J Mol Biol **342:**307-20.

156. **Srivatsan, A., Y. Han, J. Peng, A. K. Tehranchi, R. Gibbs, J. D. Wang, and R. Chen.** 2008. High-precision, whole-genome sequencing of laboratory strains facilitates genetic studies. PLoS Genet **4:**e1000139.

157. **Stetter, K. O.** 2006. Hyperthermophiles in the history of life. Philos Trans R Soc Lond B Biol Sci **361:**1837-42; discussion 1842-3.

158. **Story, S. V., C. Shah, F. E. Jenney, Jr., and M. W. Adams.** 2005. Characterization of a novel zinc-containing, lysine-specific aminopeptidase from the hyperthermophilic archaeon Pyrococcus furiosus. J Bacteriol **187:**2077-83.

159. **Szpunar, J.** 2005. Advances in analytical methodology for bioinorganic speciation analysis: metallomics, metalloproteomics and heteroatom-tagged proteomics and metabolomics. Analyst **130:**442-65.

160. **Tamames, J., R. Gil, A. Latorre, J. Pereto, F. J. Silva, and A. Moya.** 2007. The frontier between cell and organelle: genome analysis of Candidatus Carsonella ruddii. BMC Evol Biol **7:**181.

161. **Tavares, P., A. S. Pereira, J. J. Moura, and I. Moura.** 2006. Metalloenzymes of the denitrification pathway. J Inorg Biochem **100:**2087-100.

162. **Thomas-Keprta, K. L., S. J. Clemett, D. A. Bazylinski, J. L. Kirschvink, D. S. McKay, S. J. Wentworth, H. Vali, E. K. Gibson Jr, Jr., and C. S. Romanek.** 2002. Magnetofossils from ancient Mars: a robust biosignature in the martian meteorite ALH84001. Appl Environ Microbiol **68:**3663-72.

163. **Thompson, R. A., and G. R. Helz.** 1994. Copper speciation in sulfidic solutions at low sulfur activity: further evidence for cluster complexes? Geochim. Cosmochim. Acta **58:**2971-83.

164. **Tottey, S., K. J. Waldron, S. J. Firbank, B. Reale, C. Bessant, K. Sato, T. R. Cheek, J. Gray, M. J. Banfield, C. Dennison, and N. J. Robinson.** 2008. Protein-folding location can regulate manganese-binding versus copper- or zinc-binding. Nature **455:**1138-42.

165. **Trauger, S. A., E. Kalisak, J. Kalisiak, H. Morita, M. V. Weinberg, A. L. Menon, F. L. Poole, 2nd, M. W. Adams, and G. Siuzdak.** 2008. Correlating the transcriptome, proteome, and metabolome in the environmental adaptation of a hyperthermophile. J Proteome Res **7:**1027-35.

166. **Treangen, T. J., and S. L. Salzberg.** 2012. Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat Rev Genet **13:**36-46.

167. **Vaccaro, B. J., A. L. Menon, W. A. Lancaster, and M. W. Adams.** 2012. Metallomics using Inductively Coupled Plasma Mass Spectrometry. Current Protocols in Chemical Biology **(in press)**.

168. **Vaudel, M., J. M. Burkhart, D. Breiter, R. P. Zahedi, A. Sickmann, and L. Martens.** 2012. A complex standard for protein identification, designed by evolution. J Proteome Res.

169. **Verhaart, M. R., A. A. Bielen, J. van der Oost, A. J. Stams, and S. W. Kengen.** 2010. Hydrogen production by hyperthermophilic and extremely thermophilic bacteria and archaea: mechanisms for reductant disposal. Environ Technol **31:**993-1003.

170. **Waldron, K. J., and N. J. Robinson.** 2009. How do bacterial cells ensure that metalloproteins get the correct metal? Nat Rev Microbiol **7:**25-35.

171. **Waldron, K. J., J. C. Rutherford, D. Ford, and N. J. Robinson.** 2009. Metalloproteins and metal sensing. Nature **460:**823-30.

172. **Walker, C. B., Z. He, Z. K. Yang, J. A. Ringbauer, Jr., Q. He, J. Zhou, G. Voordouw, J. D. Wall, A. P. Arkin, T. C. Hazen, S. Stolyar, and D. A. Stahl.** 2009. The electron transfer system of syntrophically grown *Desulfovibrio vulgaris*. J Bacteriol **191:**5793-801.

173. **Wall, J. D., and L. R. Krumholz.** 2006. Uranium reduction. Annu Rev Microbiol **60:**149-66.

174. **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.** 2005. Protein production and crystallization at SECSG -- an overview. J Struct Funct Genomics **6:**233-43.

175. **Watanabe, S., T. Arai, R. Matsumi, H. Atomi, T. Imanaka, and K. Miki.** 2009. Crystal structure of HypA, a nickel-binding metallochaperone for [NiFe] hydrogenase maturation. J Mol Biol **394:**448-59.

176. **Waters, E., M. J. Hohn, I. Ahel, D. E. Graham, M. D. Adams, M. Barnstead, K. Y. Beeson, L. Bibbs, R. Bolanos, M. Keller, K. Kretz, X. Lin, E. Mathur, J. Ni, M. Podar, T. Richardson, G. G. Sutton, M. Simon, D. Soll, K. O. Stetter, J. M. Short, and M. Noordewier.** 2003. The genome of Nanoarchaeum equitans: insights into early archaeal evolution and derived parasitism. Proc Natl Acad Sci U S A **100:**12984-8.

177. **White, J. R., P. Escobar-Paramo, E. F. Mongodin, K. E. Nelson, and J. DiRuggiero.** 2008. Extensive genome rearrangements and multiple horizontal gene transfers in a population of *Pyrococcus* isolates from Vulcano Island, Italy. Appl Environ Microbiol **74:**6447-51.

178. **Woese, C. R., O. Kandler, and M. L. Wheelis.** 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A **87:**4576-9.

179. **Wysocki, R., and M. J. Tamas.** 2010. How Saccharomyces cerevisiae copes with toxic metals and metalloids. FEMS Microbiol Rev **34:**925-51.

180. **Xiao, Z., and A. G. Wedd.** 2010. The challenges of determining metal-protein affinities. Nat Prod Rep **27:**768-89.

181. **Yamaguchi, A., K. Iida, N. Matsui, S. Tomoda, K. Yura, and M. Go.** 2004. Het-PDB Navi.: a database for protein-small molecule interactions. J Biochem **135:**79-84.

182. **Yang, S. J., B. C. Min, Y. W. Kim, S. M. Jang, B. H. Lee, and K. H. Park.** 2007. Changes in the catalytic properties of *Pyrococcus furiosus* thermostable amylase by mutagenesis of the substrate binding sites. Appl Environ Microbiol **73:**5607-12.

183. **Yoon, S. H., D. J. Reiss, J. C. Bare, D. Tenenbaum, M. Pan, J. Slagel, R. L. Moritz, S. Lim, M. Hackett, A. L. Menon, M. W. Adams, A. Barnebey, S. M. Yannone, J. A. Leigh, and N. S. Baliga.** 2011. Parallel evolution of transcriptome architecture during genome reorganization. Genome Res **21:**1892-904.

184. **Zdobnov, E. M., and R. Apweiler.** 2001. InterProScan--an integration platform for the signature-recognition methods in InterPro. Bioinformatics **17:**847-8.

185. **Zhang, B., and S. Horvath.** 2005. A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol **4:**Article17.

186. **Zhang, W., D. E. Culley, L. Nie, and J. C. Scholten.** 2007. Comparative transcriptome analysis of *Desulfovibrio vulgaris* grown in planktonic culture and mature biofilm on a steel surface. Appl Microbiol Biotechnol **76:**447-57.

187. **Zhang, Y., and V. N. Gladyshev.** 2011. Comparative genomics of trace element dependence in biology. J Biol Chem **286:**23623-9.

188. **Zhu, W., J. W. Smith, and C. M. Huang.** 2010. Mass spectrometry-based label-free quantitative proteomics. J Biomed Biotechnol **2010:**840518.

189. **Zillig, W., I. Holz, H.-P. Klenk, J. Trent, S. Wunderl, D. Janekovic, E. Imsel, and B. Haas.** 1987. *Pyrococcus woesei*, sp. nov., an ultra-thermophilic marine archaebacterium, representing a novel order, Thermococcales. Systematic and Applied Microbiology **9:**62-70.

190. **Zivanovic, Y., P. Lopez, H. Philippe, and P. Forterre.** 2002. *Pyrococcus* genome comparison evidences chromosome shuffling-driven evolution. Nucleic Acids Res **30:**1902-10.

APPENDIX A[4]

GENOME SEQUENCING OF A GENETICALLY-TRACTABLE *PYROCOCCUS FURIOSUS* STRAIN REVEALS A HIGHLY DYNAMIC GENOME

[4] Stephanie L. Bridger*, W Andrew Lancaster*, Farris L Poole, II, Gerrit J. Schut, and Michael W. W. Adams. J Bacteriol. DOI: 10.1128/JB.00439-12
*Contributed equally to this work.

ABSTRACT

The model archaeon *Pyrococcus furiosus* grows optimally near 100°C on carbohydrates and peptides. Its genome sequence (NCBI) was determined twelve years ago. A genetically-tractable strain, COM1, was very recently reported and herein we describe its genome sequence. Of 1,910,278 bp in size, it is 2,013 bp longer (0.1%) than the reference NCBI sequence. The COM1 genome contains numerous chromosomal rearrangements, deletions and single base changes. COM1 also has 45 full or partial insertion elements compared to 35 in the reference NCBI strain and these have resulted in the direct deletion or insertional inactivation of thirteen genes. Another seven genes were affected by chromosomal deletions and are predicted to be non-functional. In addition, the amino acid sequences of another 102 of the 2134 predicted gene products are different in COM1. These changes potentially impacted various cellular functions, including carbohydrate, peptide and nucleotide metabolism, DNA repair, CRISPR-associated defense, transcriptional regulation, membrane transport and growth at 72°C. Hence, the IS-mediated inactivation of riboflavin synthase in COM1 resulted in a riboflavin-requirement for growth. Nevertheless, COM1 grew on cellobiose, malto-oligosaccharides and peptides in complex and minimal media at 98 and 72°C to the same extent as both its parent strain and a new culture collection strain (DSMZ 3638). This was in spite of COM1 lacking several metabolic enzymes, including non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase and beta-glucosidase. The *P. furiosus* genome is therefore of high plasticity and the availability of the COM1 sequence will be critical for the future studies of this model hyperthermophile.

INTRODUCTION

The genus *Pyrococcus* represents a group of obligate anaerobic archaea that grow optimally near 100°C and utilize a wide range of poly- and oligosaccharides and peptides (36). They are found in marine hydrothermal vents and are some of the most well-studied archaea with potential biotechnological applications (2). Complete genome sequences are available for five *Pyrococcus* species (9, 26, 28, 34, 49). Their evolution and genomic diversity has been linked to a high degree of DNA recombination efficiency and the presence of mobile genetic elements (58). These insertion sequences (ISs) are small (<2.5 kb), self-directed segments of DNA capable of inserting at many non-homologous sites in the target DNA (38). Numerous reports have indicated that IS element transposition has led to extensive chromosomal rearrangements and lateral gene transfer (LGT) in the *Pyrococcus* genus (13, 16, 22, 27, 58, 60). For example, environmental isolates of *Pyrococcus* differ in their distribution of IS elements suggesting a high degree of mobility (58), as has been reported for other hyperthermophilic archaea from freshwater vents (40). These genomic features are thought to play a role in the adaptation to rapidly changing environmental conditions (58).

The best characterized of the *Pyrococcus* species is *P. furiosus,* the first to be isolated (19) and the first to have its genome sequenced (49). This organism has been studied by numerous '-omics' based approaches, including transcriptomics by tiling (59) and DNA microarrays (14, 35, 51, 54, 57), comparative genomics (8, 33, 60), proteomics (32, 42), metallomics (11), and structural genomics (25). In many ways it has become one of the model hyperthermophiles, a status recently sustained by the development of a genetic system for the organism (37). This was based on isolation of a variant in a lab strain population, designated COM1, which was highly efficient in taking up and recombining exogenous DNA in both

circular and linear forms. The COM1 strain was obtained by targeted gene disruption of the *pyrF* locus (PF1114) using a plasmid designed for double-crossover recombination (37).

The COM1 strain has now been used to generate marked or markerless deletions of a number of genes, either singly or iteratively in the same strain, including those encoding two hydrogenases, a sulfur-reducing system and an iron-sulfur biosynthesis system (4, 37). A stable replicating shuttle vector is also available (18). In addition, recombinant forms have been generated that contain inserted DNA, such as highly-active promoters or encoding affinity tags for rapid protein purification (7, 24). For example, an affinity-tagged, catalytically-active, H_2 gas producing sub-complex of *P. furiosus* hydrogenase was produced at approximately 100 times the level of the native hydrogenase (24). A more robust genetic method has also been developed that allows for selection on complex growth media for agmatine auxotrophy, involving a thermostable compound required for polyamine biosynthesis (50). This was used to produce an affinity-tagged (Strep) version of the native form of the cytoplasmic hydrogenase of *P. furiosus* (7).

The development of additional genetic tools for COM1 and its use as a platform for genetic manipulation and metabolic engineering can be anticipated. For such studies the complete genome sequence of this strain is obviously essential for all molecular manipulations. It was also important to determine if the genome sequence of COM1 could reveal any unanticipated phenotypic changes, particularly those affecting its ability to metabolize growth substrates or its requirement for vitamins or cofactors. The genome sequence of COM1 has now been determined and comparison with the published sequence of the *P. furiosus* reference strain ((49); NC_003413) revealed a surprisingly large number of changes. The results of this study have implications not only for the utility of this new genetic system for a well-studied

hyperthermophile but also for our understanding of the dynamics of genomic change and chromosome maintenance in *P. furiosus* and of the function of two carbohydrate metabolism enzymes, GAPN and CelB.

MATERIALS & METHODS

Strains and growth conditions.

The three *P. furiosus* strains used in this study are termed COM1, which is the genetically tractable strain (37); Parent, which is the lab strain from which COM1 was derived, and DSMZ control strain (3638), which was acquired in October 2010 from the German Collection of Microorganisms and Cell Cultures (Braunshweig, Germany). For genomic DNA isolation (and routine growth studies), cells were grown from a single colony of the COM1 strain, the Parent lab strain population, and the minimally passaged DSMZ control strain. They were grown in a complex cellobiose medium as described previously (37) with uracil (20 μM) added to the growth medium of the auxotrophic COM1 strain. Growth experiments to evaluate phenotypes were carried out in biological triplicate in 100 mL serum bottles with 50 mL of complex or defined media with cellobiose, maltose or malto-oligosaccarides as carbon sources at 98°C or 72°C. To obtain cells for enzyme assays, cultures were grown on cellobiose and peptides in a 20 L DCI-Biolafitte BioPro^{TM} Evo Series SIP fermentor and cells were harvested at the end of exponential growth.

DNA sequencing.

Cells for genomic DNA isolation were harvested from a 400 mL stationary phase culture and suspended in 2 mL buffer A (25% sucrose, 50 mM Tris-HCl, 40 mM EDTA pH 7.4) followed by incubation at 37°C for 1 hr with 0.6 mg/mL RNase, 0.2 mg/mL Proteinase K, and

0.25 M EDTA pH 8.0, and then incubation at 65°C for 45 min with 1% SDS, Triton X-100, cetyltrimethylammonium bromide (CTAB), and 0.7M NaCl. Genomic DNA was extracted using phenol:chloroform:isoamyl alcohol (25:24:1, buffered at pH 8), ethanol precipitated and suspended in 10 mM Tris buffer, pH 8.0, to a concentration of ~0.1µg/µl. All genomic sequencing libraries were prepared according to the manufacturer's guidelines. Single-end 454 sequencing of COM1 was performed using a Roche GS FLX titanium pyrosequencer, with a quarter-plate. Illumina sequencing was performed using a single lane on a GAIIx sequencer. The mate-pair library was constructed with a 2-5 kb protocol. Sanger sequencing of PCR products was used to confirm mutations and genome rearrangements. PCR targets were amplified using PfuTurbo® DNA Polymerase (Stratagene) on a Bio-Rad C1000™ thermal cycler. Products were analyzed on an agarose gel and purified using a StrataPrep PCR Purification Kit (Agilent Technologies).

Genome sequences, assembly and analysis.

De novo hybrid assembly of COM1 single-end 454 and mate-pair Illumina data was performed using the MIRA assembler (version 3.2.1). Illumina library insert size was estimated using Bowtie version 0.12.7 on the NCBI reference *P. furiosus* sequence (NC_003413). Scaffolding of the resulting MIRA-produced contigs was performed using SSPACE version 1.1 in extension mode with the mate-pair Illumina reads. Scaffold regions with ambiguous base calls were PCR amplified and sequenced. Genome block alignments were performed using Progressive Mauve version 2.3.1. Fuzznuc, part of the Emboss toolkit 6.3.1 (48), was used to locate ORFs with exact matches to NCBI-annotated sequences and to sequences derived from a mapped assembly using the *P. furiosus* NCBI genome as a reference. Additional ORFs present

in the NCBI annotation that did not have exact matches in the *de novo* assembled chromosome were located using BLAST version 2.2.24 and coordinates determined by visual inspection. In addition to the NCBI gene names, Interpro's Iprscan version 4.8 was used to assign functional annotations to ORFs. Transmembrane domains and signal peptides were predicted using the optional TMHMM and SignalP packages. IS elements were determined using the ISFinder BLAST analysis tool (http://www-is.biotoul.fr/; (52)). The COM1 assembled genome was visualized and annotations corrected using the CLC Genomics Workbench version 6.2. Annotated ORFs were translated using Transeq (EMBOSS toolkit 6.3.1) with Table 11. Needleman-Wunsch global alignments of the translated ORFs with NCBI reference protein sequences were performed using Needle (EMBOSS toolkit 6.3.1). An alignment of the NCBI reference genome (NC_003413) with COM1 was performed using Nucmer version 3.0.7 with a maximum gap size of 500 and a minimum cluster length of 100. A dot plot of the alignment was created using Mummerplot version 3.5.

Enzyme activities.

Cytoplasmic extracts of each strain were prepared as previously described (42). All steps were carried out under strict anaerobic conditions unless otherwise noted. Non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase (GAPN) activity was measured at 70°C monitoring the rate of NAD(P)H formation at 340 nm (41). β-glucosidase (CelB) activity was measured aerobically in a discontinuous assay with β-D-glucopyranoside-*p*-nitrophenyl as the substrate at 80°C (29). Reaction mixtures contained 50 mM Tris-HCl (pH 7.4) for GAPN and 50 mM sodium phosphate (pH 6.0) for CelB. The Bradford method (3) was used to estimate protein concentrations in cell extracts using bovine serum albumin as the standard.

RESULTS

Genome sequence.

De novo assembly of 5,213,746 total reads of the COM1 genomic DNA resulted in one scaffold of 1,910,278 bp. This is 2,013 bp longer (0.1%) than the NCBI reference sequence for *P. furiosus* (1,908,256 bp) (49). The average coverage of the COM1 scaffold was 32X and 89X for 454 and Illumina reads, respectively. The fully assembled DNA sequence of COM1 has been deposited in GenBank (accession number AKVM00000000.1).

Genome organization.

Alignment of the COM1 genome sequence with the *P. furiosus* NCBI reference (49) revealed a high overall degree of synteny, with two major inversions (Fig. 1). Sequence blocks, labeled A to E, were assigned to the COM1 genome based on the organization of the reference sequence (Figs. 2 A and B, respectively) with block A (red colored arrow) starting at PF0001. Block boundaries are located at the sites of IS elements (black tick marks), with the exception of the boundary between blocks A and E, which represents the boundary between the first and last annotated ORFs in the NCBI reference sequence (PF0001 and PF2065). Block A (red) begins at position 1,497,249 in the COM1 genome sequence and continues through PF0189, followed by block D (blue, PF0390 – PF1603'), block C (green, PF0349 – PF0388), block B (yellow, PF0190 - PF0347), and block E (purple, PF1603'' – PF2065). Relative to the NCBI reference sequence, blocks AE and C are inverted in COM1. The IS elements identified at the boundaries flanking block C were part of the originally annotated sequence (PF0348 and PF0389), whereas the other sites of rearrangement are due to recent IS activity in the intergenic region upstream of PF0190 and within PF1603.

Confirmation of the COM1 genome organization (Figs. 1 and 2A) was obtained by amplifying and sequencing PCR products that spanned the boundaries of each genome block (see Fig. S1 and S2 in the supplemental material). Care was taken to design unique primers not within transposable elements. The chromosomal orientations of the Parent and DSMZ strains were also analyzed. Surprisingly, both the NCBI reference order (Fig. 2B) and the inverted orientations of blocks AE and C were observed in the COM1 and Parent. Only the inverted order of block C and the reference order of block AE were observed in the DSMZ control strain.

IS element activity.

COM1 and the *P. furiosus* reference sequences were annotated for IS elements using ISFinder blast analysis tool (52) with an e-value cut off of $\leq$ 4E-17. The reference sequence contains 29 complete and 6 partial IS elements (Table 1), whereas COM1 also contains the same 6 partial sequences in the same relative positions, but the number of complete sequences increased to 39. IS elements represent over 26 kb (1.38%) of the reference genome but over 34 kb (1.79%) of the COM1 genome. Most of the IS elements (80%) identified in both genomes are members of the IS6 family with the remainder representing three other IS families (IS982, IS607 and IS200/IS605, Table 1). Species-specific names for the IS elements are based on their origin and include not only ISPfu1-ISPfu5 but also those derived from *Thermococcus* species (ISTko2 and ISTsi3, (52)). The full-length copies of the different ISPfu elements (1-5) have consistent sizes (781, 782, 933, 1961, and 779 bp, respectively) and possess properties typical of known transposable elements, i.e., putative transposase genes, terminal inverted repeats, and flanking duplicated target sequences. Locations of these IS elements were mapped (Figs. 2A and B) and it is evident that they are responsible for a large number of genomic changes in COM1 relative to

the NCBI sequence. These are summarized in Table 2. They include four excisions resulting in the complete or partial deletion of six neighboring genes, seven insertions within seven genes, and nine intergenic insertions potentially affecting regulatory regions upstream of eleven genes. Of these 24 genes, 10 encode conserved hypothetical proteins, 13 are predicted to be in operons, and the predicted cellular location of the gene products are divided between the membrane (7), extracellular (5) and cytoplasmic fractions (12).

Transposases originally annotated in the NCBI reference sequence as PF0013 (ISPfu1), PF0408 (ISPfu1), PF2035 (ISPfu2) and PF0756 (ISPfu2) are deleted in COM1, and these include portions of or complete neighboring genes: PF0012 (39 bp 3' end), PF0407 (41 bp 5' end), PF2034 (41 bp 5' end), and a 1,714 bp region with PF0755 (300 bp deleted at 3' end), PF0757 and PF0758 (deleted), respectively (Table 2). The affected neighboring genes are predicted to be non-functional since, in most cases, the deleted region includes the transcriptional start site. Additionally, IS-mediated gene disruptions include PF0061, PF0393, PF0429, PF0823, PF1260, PF1603 and PF2059 and these are also predicted to be non-functional. Nucleotide alignments of each gene are identical to the NCBI reference sequence minus the IS element and duplicated target repeat. However, at the protein-level, the COM1 sequences terminate with premature stop codons ~16 bp into the IS element. An example of IS element gene disruption is shown in Fig. 3, which compares the gene encoding riboflavin synthase subunit alpha (PF0061). There are some other IS-disrupted genes that have a specific rather than a general predicted function, and these include GAPN, aminopeptidase, and proline permease (Table 2), and these are considered further below. Similarly, several intergenic regions were also disrupted by IS elements and may potentially affect the regulation of genes with specific known or predicted functions, which

include alpha-amylase, sugar kinase, trehalose/maltose binding protein, dihydroorotase, methyltransferase, purine permease and cold-induced protein A (57), as shown in Fig. 4.

In addition to the IS-mediated excisions observed in COM1, there are several genes affected by large chromosomal deletions compared to the reference sequence that appear to be independent of these elements. However, given the obvious recent IS activity, it is impossible to rule out previous IS activity in these regions prior to excision. The coding sequences of the seven genes are partially or completely deleted (Table 3) and the encoded proteins are therefore predicted to be non-functional. These include CelB (PF0073), which has a 774 bp deletion at the 3' end resulting in protein length of 214 compared to 472 amino acid residues in the NCBI reference sequence. The other six genes are in a single large chromosomal region (6,238 bp) that includes the genes PF1249-PF1253 (deleted) and PF1254 (904 bp deleted at 5' end). These encode an ABC transporter, a sodium dependent transporter, aromatic aminotransferase II and three hypothetical proteins.

Protein-level genome differences.

In addition to the IS-mediated gene disruptions and large chromosomal deletions observed in COM1, the protein products of 102 of the total 2134 genes in the NCBI reference sequence are disrupted in COM1 at the translational level through non-synonymous mutations. These result in amino acid changes, insertions and deletions introducing premature stop codons and frameshifts producing longer or shorter products with alternate coding sequences. These were subdivided into 36 major changes (Table 4) and 66 minor changes (see Table S1 in supplemental material) based on the Needleman-Wunsch global alignment identity when compared with the protein sequences in the NCBI reference sequence (49) using $\geq$ 90% cutoff as

a major difference. Since the deletion of the *pyrF* gene (PF1114) is the auxotrophic selectable marker for uracil biosynthesis in the COM1 strain (37), it was omitted from the list of major differences in Table 4 but is included in the complete list of disrupted genes in COM1 relative to the NCBI reference (see Table S1 in supplemental material).

In general, although the COM1 genes that fall into the major difference category almost certainly encode nonfunctional proteins, those in the minor difference category could also result in protein products with diminished or abolished functions. Experimental assessment of function would be required for proteins deemed to have minor changes since a single amino acid change could obviously lead to inactivation. Functional annotation of all 102 genes in the reference sequence revealed that 46 of them encode conserved hypothetical proteins (see Table S1 in supplemental material). The remaining 56 are predicted to be involved in a wide variety of cellular functions involving carbohydrate and peptide metabolism, DNA repair, CRISPR-associated defense, transcriptional regulation and membrane transport. However, as described above for the IS-mediated disruptions, there are only some affected genes where a specific rather than a general function can be assigned. For example, the major difference category includes two enzymes involved in carbohydrate degradation, β-galactosidase (PF0363) and amylopullulanase (PF1935*; Table 4). β-galactosidase in COM1 is predicted to be non-functional as it contains only 32 of the 772 amino acid residues encoded in the NCBI reference sequence due to a single base deletion only 18 bp into the coding sequence creating a premature stop. On the other hand, the gene encoding amylopullulanase was previously reported to have a sequencing error that revealed a gene fusion of PF1934-PF1935 (35). In COM1 we found that the gene fusion results in an even longer protein than previously reported (1,355 compared to 1,114 amino acids, (35)), and is more similar in length to those in other *Pyrococcus* and

*Thermococcu*s species. The amylopullulanase in COM1 is therefore expected to be fully functional. In fact, PCR confirmed the same length product for the PF1935* gene in all three strains (DSMZ, Parent, and COM1, data not shown). The COM1 genome sequence also contained the gene fusion between PF1191 and PF1192 previously demonstrated by the native purification of spherical virus-like particles from *P. furiosus* (44).

Phenotypic properties.

The COM1, Parent and DSMZ strains were screened for their ability to grow under various conditions to investigate the predictions of non-functional genes based on the COM1 sequence. For example, both IS and non-IS disruption of gene products or gene regulatory regions (Tables 2-4) involve β-glucosidase (CelB; PF0073), 4-α-glucanotransferase (PF0272), sugar kinase (PF1738), trehalose/maltose binding protein (PF1739) and GAPN (PF0755). However the growth rates and final cell densities of all three strains, COM1, Parent, and DSMZ, on cellobiose, maltose and malto-oligosaccharides were comparable, with the proviso for COM1 that yeast extract (0.5 g/L) was also added. This was due to the inability of COM1 to synthesize riboflavin due to a non-functional riboflavin synthase (PF0061) (Table 2). When riboflavin (0.1 μM) was added, the yeast extract was no longer required. COM1 also grew on a complex peptide-based medium (tryptone, 5 g/L) in the presence of elemental sulfur (2 g/L) to the same extent as the Parent and DSMZ strains, indicating that the disrupted genes in COM1 potentially related to peptide catabolism, including aromatic aminotransferase II (PF1253), proline permease (PF0429) and aminopeptidase (PF2059) (Tables 2 and 3), do not affect growth on peptides. In addition, all three strains grew well on the defined minimal medium (although COM1 required riboflavin) implying that the function of gene products involved in nucleotide metabolism, such

as dihydroorotase (PF0189) and purine permease (PF1240) (Table 2), are still fully functional despite upstream IS insertion or not required for growth under these conditions.

A previous study (57) showed that a number of genes are up-regulated when *P. furiosus* is grown at 72°C compared to the optimum near 100°C, the most highly-expressed of which is a membrane glycoprotein termed cold-induced protein A (CipA, (PF0190)). In COM1, *cipA* is not expressed due to the upstream insertion of an IS element (Table 2 and Fig. 4). However, no differences were observed in the ability of the COM1, Parent and DSMZ strains in their abilities to grow at low temperature (72°C, data not shown). In addition, a number of genes involved in DNA repair, including an uncharacterized RadA domain protein (PF0872), DNA repair helicase Rad3 (PF0933), a 5' to 3' exonuclease NurA (PF1168, (23)), and DNA repair helicase (PF1902) could potentially be inactive in COM1 as they fall into the minor protein-level differences category (see Table S1 in supplemental material). However, COM1, Parent and DSMZ showed no significant differences in their ability to recover from exposure to UV or gamma irradiation using conditions previously reported for *P. furiosus* (14). Consequently, in spite of the massive genome rearrangements in COM1 relative to the NCBI reference sequence, and the actual or potential inactivation of more than 120 genes, its ability to grow under various conditions previously used for the Parent strain in the laboratory are unaffected.

Enzyme assays were carried out to determine the effects of 5' end deletion of the gene encoding non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase, GAPN (PF0755), and the 3' end deletion of the gene encoding β-glucosidase (CelB; PF0073). The DSMZ strain was used as a control, as PCR and sequence analyses confirmed that it contained full-length versions of *gapn* and *celB* (data not shown). On the other hand, PCR analysis of the Parent strain revealed a full-length copy of *gapn* but its *celB* contained the same 5' end deletion as that found

in COM1. Surprisingly, all attempts to measure GAPN activity in the cell extracts of all three of the strains were unsuccessful, even in the presence of glucose-1-phosphate (1 mM), which is reportedly an activator of the enzyme in *Thermococcus kodakarensis* (41). As expected, the DSMZ control strain contained high CelB activity (1.68 µmoles/min/mg at pH 6.0) and this was an order of magnitude greater than that measured in extracts of either the COM1 or Parent strains (0.16 and 0.15 µmoles/min/mg, respectively at pH 6.0).

DISCUSSION

The availability of a highly efficient genetically-tractable strain of *P. furiosus* (COM1) has dramatically expanded the potential to study this model hyperthermophilic species. Its sequence not only defines the genetic platform for these future studies, it has also provided novel insights into the biology of this organism including the dynamic nature of its genome and the utility of two classical carbohydrate metabolism genes, *gapn* and *celB*. The COM1 sequence assembly and scaffolding indicated the inversion of two large chromosomal segments relative to the published NCBI reference sequence (Figs. 1 and 2) and PCR confirmation revealed that both the reference and inverted order of these segments were present in genomic DNA isolated from a small scale culture originating from a single colony. This remarkable flexibility of chromosomal arrangement without apparent consequence to proper metabolic function has not previously been observed in this organism. The high rate of chromosomal breakage in hyperthermophilic environments and those where large doses of ionizing radiation are encountered require robust and accurate repair mechanisms (46). *P. furiosus* is able to completely reassemble its chromosome after gamma radiation-induced fragmentation into 30-500 kb fragments with up to 15 chromosomal copies per cell present in the exponential growth phase serving as templates for reassembly (12). This ability, coupled with IS elements scattered throughout the genome that

serve as substrates for homologous recombination, is consistent with the observed alterations in genome structure in COM1. A comparative genomics study of three *Pyrococcus* species revealed that the genome of *P. furiosus* exhibits less bias in the preference of co-directionality of transcription with replication than other sequenced organisms at the time (60). Given the ease with which the genome undergoes rearrangement, this is likely necessary to ensure proper transcription of essential genes. In addition, the high degree of operonization in the genome of *P. furiosus* places further constraints on the location of chromosomal shuffling hot spots (59).

IS elements also participate in genome evolution by disrupting gene coding sequences, influencing the expression of genes downstream of insertions (10). Sequencing of the COM1 genome revealed that two IS element types, ISPfu1 and 2 (both of the IS6 family), have been recently active under laboratory conditions and involved in four excisions and fourteen insertions within genes and upstream of genes in potential regulatory regions (Table 2). LGT of a bacterial-type composite transposon was previously documented in *P. furiosus* and *T. litoralis* with the acquisition in the former organism of the Mal I ABC transport system for maltose and trehalose (13, 22, 45). Among the seven IS annotated archaeal genomes (30), *P. furiosus* harbors far more IS elements (29 full, 6 partial) than any other member of the *Thermococcales*, but less than a quarter of the IS elements identified in the *S. solfataricus* genome (146 full, 297 partial). In fact, no full-length IS elements have been identified in the genomes of *P. abyssi* or *P. horikoshii*, and only one full length IS element has been identified in *T. kodakarensis* (30).

In addition to IS-mediated disruptions, 102 differences in sequence were observed in the COM1 genome that would lead to gene products differing from those encoded by the published NCBI reference sequence. Some are likely to have been sequencing or assembly errors in the original published NCBI sequence, some could have arisen in the lab strain through random

mutagenesis and some may be unique to the COM1 strain. However, these categories are difficult to assess quantitatively. Although the NCBI reference sequence is labeled as the type strain (DSM 3638), the strain that provided the reference sequence (49) was not obtained directly from the DSMZ (Kelly, R. and Robb, F., personal communication). Originally a gift from Dr. Karl Stetter (19), the strain was maintained at 4°C in a laboratory setting with periodic clonal selection for more than 10 years prior to producing DNA for genome sequencing in 1999 (49). Similarly, the Parent strain that gave rise to COM1 (37) was maintained in the laboratory for two years at ambient temperature with multiple transfers on various media after being obtained from DSMZ in 2007. Therefore, it is assumed that the DSMZ control strain (2010) used in this study is a good representative of the ancestor of Parent and COM1. That maintenance under laboratory conditions can lead to numerous mutations in strains originating from the same isolate is not surprising and is now well established in other organisms. For example, this was recently revealed by a large scale re-sequencing effort of *Bacillus subtilis* strains from multiple laboratories (53), emphasizing the need to maintain a permanent storage capacity within the laboratory setting.

Despite extensive genomic changes in the COM1 strain, including multiple excisions, insertions and protein-level changes, its major metabolic functions have not been disrupted. Select phenotypic properties were examined based on the changes in the COM1 strain but no significant differences in growth under a variety of conditions were detected between it and the Parent and DSMZ strains. However, these analyses did reveal surprisingly non-essential roles for two carbohydrate metabolism enzymes, GAPN and CelB, whose functions should be abolished in COM1 due to large deletions in their coding regions. GAPN has been shown to be a key glycolytic enzyme in the related euryarchaeon *T. kodakarensis* since a strain with a targeted

gene knockout (Δ*gapN*) was unable to grow on malto-oligosaccharides (41). Similarly, GAPN plays a key glycolytic role in various crenarchaeota, including *Thermoproteus* (5), *Aeropyrum* (47) and *Sulfolobus* species (1, 17). The oxidation of glyceraldehyde-3-phosphate during glycolysis in *P. furiosus* can be catalyzed by the tungsten-containing enzyme, glyceraldehyde-3-phosphate ferredoxin oxidoreductase (GAPOR), which uses ferredoxin rather than NADP as the electron acceptor (43). In *T. kodakarensis* both GAPOR and GAPN are required for growth on sugars (41) but, clearly this is not the case in *P. furiosus,* as COM1 grew well on maltose, cellobiose and malto-oligosaccharides in the absence of a functional GAPN. COM1 also lacks the gene encoding CelB, which has been shown to be the cellobiose hydrolyzing enzyme in *P. furiosus* that constitutes up to 5% of the total cell protein during growth on this β-1,4-linked disaccharide (29). An order of magnitude less CelB activity was measured in COM1 cell extracts compared to that of the DSMZ control strain harboring a full-length *celB g*ene. Therefore, growth of COM1 on cellobiose must be accomplished via other beta-specific glycosidases present in *P. furiosus* (15).

One significant phenotypic difference between COM1 and the Parent and DSMZ strains is obviously its ability to be genetically manipulated. The high efficiency in transformation and recombination of COM1 is likely the result of changes in multiple genes but genome sequence comparisons between the strains cannot define the responsible gene(s). One group potentially involved in conferring competence is the disruption of five of the CRISPR-associated (Cas) proteins (Tables 2 and 4 and see Table S1 in supplemental material). This is the most well characterized defense system against invasion by foreign nucleic acids in both archaea and bacteria (20, 39). In *P. furiosus*, Cas6 has been shown to recognize and cleave foreign RNA that match spacer sequences present in the host's CRISPR loci (6, 21). A possible role of these

disrupted proteins in conferring competence is intriguing and an obvious target for future research. Other disrupted genes in COM1 that could also play a role in DNA uptake and recombination include 20 membrane transporters, one or more of which could affect membrane integrity thereby facilitating DNA uptake. Similarly, numerous proteins involved in DNA replication, transcription, and repair are actually or potentially disrupted in COM1 (see Tables 2 and 4). These include four DNA helicases (PF0085, PF0572, PF0933, PF1902), two RNA helicases (PF0592, PF1120), a DNA primase (PF1725), a cell division control protein (PF1882), an uncharacterized RadA domain protein (PF0872), five putative transcription factors (DNA-binding proteins: PF0054, PF0524, PF0621, PF0710, PF1206), a RNA-binding protein (PF1573), and a 5' to 3' exonuclease NurA (PF1168, (23)). In addition, of the 122 genes disrupted by IS- and non-IS-mediated events in the COM1 strain, 56 of them encode conserved hypothetical proteins, some of which could be partially or solely responsible for the genetic properties of COM1. Clearly, determining how many of these genes directly or indirectly affect competence will not be an easy task.

ACKNOWLEDGEMENTS

We acknowledge the Division of Chemical Sciences, Geosciences, and Biosciences, Office of Basic Energy Sciences of the U.S. Department of Energy through grant DE-FG05-95ER20175 for funding the phenotypic analyses and the ENIGMA project supported by the U. S Department of Energy under Contract No. DE-AC02-05CH11231 for funding the sequencing and data analysis. We thank the UGA Sequencing Facility for 454 sequencing, K. Stirrett and J. Westpheling for providing a sample of COM1 DNA for 454 sequencing, Ryan Weil at the

Emory GRA Genomics Core for assistance with Illumina sequencing, Mirko Basen for assistance with low temperature cultures, and Gina Lipscomb for many helpful discussions.

REFERENCES

1. **Albers, S. V., N. K. Birkeland, A. J. Driessen, S. Gertig, P. Haferkamp, H. P. Klenk, T. Kouril, A. Manica, T. K. Pham, P. Ruoff, C. Schleper, D. Schomburg, K. J. Sharkey, B. Siebers, P. Sierocinski, R. Steuer, J. van der Oost, H. V. Westerhoff, P. Wieloch, P. C. Wright, and M. Zaparty.** 2009. SulfoSYS (Sulfolobus Systems Biology): towards a silicon cell model for the central carbohydrate metabolism of the archaeon *Sulfolobus solfataricus* under temperature variation. Biochem Soc Trans **37:**58-64.

2. **Atomi, H.** 2005. Recent progress towards the application of hyperthermophiles and their enzymes. Curr Opin Chem Biol **9:**166-73.

3. **Bradford, M. M.** 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem **72:**248-54.

4. **Bridger, S. L., S. M. Clarkson, K. Stirrett, M. B. DeBarry, G. L. Lipscomb, G. J. Schut, J. Westpheling, R. A. Scott, and M. W. Adams.** 2011. Deletion strains reveal metabolic roles for key elemental sulfur-responsive proteins in *Pyrococcus furiosus*. J Bacteriol **193:**6498-504.

5. **Brunner, N. A., H. Brinkmann, B. Siebers, and R. Hensel.** 1998. NAD+-dependent glyceraldehyde-3-phosphate dehydrogenase from *Thermoproteus tenax*. The first identified archaeal member of the aldehyde dehydrogenase superfamily is a glycolytic enzyme with unusual regulatory properties. J Biol Chem **273:**6149-56.

6. **Carte, J., N. T. Pfister, M. M. Compton, R. M. Terns, and M. P. Terns.** 2010. Binding and cleavage of CRISPR RNA by Cas6. RNA **16:**2181-8.

7. **Chandrayan, S. K., P. M. McTernan, R. C. Hopkins, J. Sun, F. E. Jenney, Jr., and M. W. Adams.** 2012. Engineering Hyperthermophilic Archaeon *Pyrococcus furiosus* to Overproduce Its Cytoplasmic [NiFe]-Hydrogenase. J Biol Chem **287:**3257-64.

8. **Chinen, A., I. Uchiyama, and I. Kobayashi.** 2000. Comparison between *Pyrococcus horikoshii* and *Pyrococcus abyssi* genome sequences reveals linkage of restriction-modification genes with large genome polymorphisms. Gene **259:**109-21.

9. **Cohen, G. N., V. Barbe, D. Flament, M. Galperin, R. Heilig, O. Lecompte, O. Poch, D. Prieur, J. Querellou, R. Ripp, J. C. Thierry, J. Van der Oost, J. Weissenbach, Y.**

Zivanovic, and P. Forterre. 2003. An integrated analysis of the genome of the hyperthermophilic archaeon *Pyrococcus abyssi*. Mol Microbiol **47:**1495-512.

10. **Craig, N. L., R. Craigie, M. Gellert, and A. M. Lambowitz (ed.).** 2002. Mobile DNA II. American Society for Microbiology, Washington, D.C.

11. **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.** 2010. Microbial metalloproteomes are largely uncharacterized. Nature **466:**779-82.

12. **DiRuggiero, J., J. R. Brown, A. P. Bogert, and F. T. Robb.** 1999. DNA Repair Systems in Archaea: Mementos from the Last Universal Common Ancestor? J Mol Evol **49:**474-84.

13. **Diruggiero, J., D. Dunn, D. L. Maeder, R. Holley-Shanks, J. Chatard, R. Horlacher, F. T. Robb, W. Boos, and R. B. Weiss.** 2000. Evidence of recent lateral gene transfer among hyperthermophilic archaea. Mol Microbiol **38:**684-93.

14. **DiRuggiero, J., N. Santangelo, Z. Nackerdien, J. Ravel, and F. T. Robb.** 1997. Repair of extensive ionizing-radiation DNA damage at 95 degrees C in the hyperthermophilic archaeon *Pyrococcus furiosus*. J Bacteriol **179:**4643-5.

15. **Driskill, L. E., M. W. Bauer, and R. M. Kelly.** 1999. Synergistic interactions among beta-laminarinase, beta-1,4-glucanase, and beta-glucosidase from the hyperthermophilic archaeon *Pyrococcus furiosus* during hydrolysis of beta-1,4-, beta-1,3-, and mixed-linked polysaccharides. Biotechnol Bioeng **66:**51-60.

16. **Escobar-Paramo, P., S. Ghosh, and J. DiRuggiero.** 2005. Evidence for genetic drift in the diversification of a geographically isolated population of the hyperthermophilic archaeon *Pyrococcus*. Mol Biol Evol **22:**2297-303.

17. **Ettema, T. J., H. Ahmed, A. C. Geerling, J. van der Oost, and B. Siebers.** 2008. The non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase (GAPN) of *Sulfolobus solfataricus*: a key-enzyme of the semi-phosphorylative branch of the Entner-Doudoroff pathway. Extremophiles **12:**75-88.

18. **Farkas, J., D. Chung, M. DeBarry, M. W. Adams, and J. Westpheling.** 2011. Defining components of the chromosomal origin of replication of the hyperthermophilic

archaeon *Pyrococcus furiosus* needed for construction of a stable replicating shuttle vector. Appl Environ Microbiol **77:**6343-9.

19. **Fiala, G., and K. Stetter.** 1986. *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100°C. Arch Microbiol **145:**56-61.

20. **Garrett, R. A., G. Vestergaard, and S. A. Shah.** 2011. Archaeal CRISPR-based immune systems: exchangeable functional modules. Trends Microbiol **19:**549-56.

21. **Hale, C. R., P. Zhao, S. Olson, M. O. Duff, B. R. Graveley, L. Wells, R. M. Terns, and M. P. Terns.** 2009. RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. Cell **139:**945-56.

22. **Hamilton-Brehm, S. D., G. J. Schut, and M. W. Adams.** 2005. Metabolic and evolutionary relationships among *Pyrococcus* Species: genetic exchange within a hydrothermal vent environment. J Bacteriol **187:**7492-9.

23. **Hopkins, B. B., and T. T. Paull.** 2008. The *P. furiosus* mre11/rad50 complex promotes 5' strand resection at a DNA double-strand break. Cell **135:**250-60.

24. **Hopkins, R. C., J. Sun, F. E. Jenney, Jr., S. K. Chandrayan, P. M. McTernan, and M. W. Adams.** 2011. Homologous expression of a subcomplex of *Pyrococcus furiosus* hydrogenase that interacts with pyruvate ferredoxin oxidoreductase. PLoS One **6:**e26569, Oct 24 (epub).

25. **Jenney, F. E., Jr., and M. W. Adams.** 2008. The impact of extremophiles on structural genomics (and vice versa). Extremophiles **12:**39-50.

26. **Jun, X., L. Lupeng, X. Minjuan, P. Oger, W. Fengping, M. Jebbar, and X. Xiang.** 2011. Complete genome sequence of the obligate piezophilic hyperthermophilic archaeon *Pyrococcus yayanosii* CH1. J Bacteriol **193:**4297-8.

27. **Kanoksilapatham, W., J. M. Gonzalez, D. L. Maeder, J. DiRuggiero, and F. T. Robb.** 2004. A proposal to rename the hyperthermophile *Pyrococcus woesei* as *Pyrococcus furiosus* subsp. *woesei*. Archaea **1:**277-83.

28. **Kawarabayasi, Y., M. Sawada, H. Horikawa, Y. Haikawa, Y. Hino, S. Yamamoto, M. Sekine, S. Baba, H. Kosugi, A. Hosoyama, Y. Nagai, M. Sakai, K. Ogura, R.**

Otsuka, H. Nakazawa, M. Takamiya, Y. Ohfuku, T. Funahashi, T. Tanaka, Y. Kudoh, J. Yamazaki, N. Kushida, A. Oguchi, K. Aoki, and H. Kikuchi. 1998. Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium, *Pyrococcus horikoshii* OT3. DNA Res **5:**55-76.

29. **Kengen, S. W., E. J. Luesink, A. J. Stams, and A. J. Zehnder.** 1993. Purification and characterization of an extremely thermostable beta-glucosidase from the hyperthermophilic archaeon *Pyrococcus furiosus*. Eur J Biochem **213:**305-12.

30. **Kichenaradja, P., P. Siguier, J. Perochon, and M. Chandler.** 2010. ISbrowser: an extension of ISfinder for visualizing insertion sequences in prokaryotic genomes. Nucleic Acids Res **38:**D62-8.

31. **Koning, S. M., W. N. Konings, and A. J. Driessen.** 2002. Biochemical evidence for the presence of two alpha-glucoside ABC-transport systems in the hyperthermophilic archaeon *Pyrococcus furiosus*. Archaea **1:**19-25.

32. **Lancaster, W. A., J. L. Praissman, F. L. Poole, 2nd, A. Cvetkovic, A. L. Menon, J. W. Scott, F. E. Jenney, Jr., M. P. Thorgersen, E. Kalisiak, J. V. Apon, S. A. Trauger, G. Siuzdak, J. A. Tainer, and M. W. Adams.** 2011. A computational framework for proteome-wide pursuit and prediction of metalloproteins using ICP-MS and MS/MS data. BMC Bioinformatics **12:**64.

33. **Lecompte, O., R. Ripp, V. Puzos-Barbe, S. Duprat, R. Heilig, J. Dietrich, J. C. Thierry, and O. Poch.** 2001. Genome evolution at the genus level: comparison of three complete genomes of hyperthermophilic archaea. Genome Res **11:**981-93.

34. **Lee, H. S., S. S. Bae, M. S. Kim, K. K. Kwon, S. G. Kang, and J. H. Lee.** 2011. Complete genome sequence of hyperthermophilic *Pyrococcus* sp. strain NA2, isolated from a deep-sea hydrothermal vent area. J Bacteriol **193:**3666-7.

35. **Lee, H. S., K. R. Shockley, G. J. Schut, S. B. Conners, C. I. Montero, M. R. Johnson, C. J. Chou, S. L. Bridger, N. Wigner, S. D. Brehm, F. E. Jenney, Jr., D. A. Comfort, R. M. Kelly, and M. W. Adams.** 2006. Transcriptional and biochemical analysis of starch metabolism in the hyperthermophilic archaeon *Pyrococcus furiosus*. J Bacteriol **188:**2115-25.

36. **Leigh, J. A., S. V. Albers, H. Atomi, and T. Allers.** 2011. Model organisms for genetics in the domain Archaea: methanogens, halophiles, *Thermococcales* and *Sulfolobales*. FEMS Microbiol Rev **35:**577-608.

37. **Lipscomb, G. L., K. Stirrett, G. J. Schut, F. Yang, F. E. Jenney, Jr., R. A. Scott, M. W. Adams, and J. Westpheling.** 2011. Natural competence in the hyperthermophilic archaeon *Pyrococcus furiosus* facilitates genetic manipulation: construction of markerless deletions of genes encoding the two cytoplasmic hydrogenases. Appl Environ Microbiol **77:**2232-8.

38. **Mahillon, J., and M. Chandler.** 1998. Insertion sequences. Microbiol Mol Biol Rev **62:**725-74.

39. **Makarova, K. S., D. H. Haft, R. Barrangou, S. J. Brouns, E. Charpentier, P. Horvath, S. Moineau, F. J. Mojica, Y. I. Wolf, A. F. Yakunin, J. van der Oost, and E. V. Koonin.** 2011. Evolution and classification of the CRISPR-Cas systems. Nat Rev Microbiol **9:**467-77.

40. **Martusewitsch, E., C. W. Sensen, and C. Schleper.** 2000. High spontaneous mutation rate in the hyperthermophilic archaeon *Sulfolobus solfataricus* is mediated by transposable elements. J Bacteriol **182:**2574-81.

41. **Matsubara, K., Y. Yokooji, H. Atomi, and T. Imanaka.** 2011. Biochemical and genetic characterization of the three metabolic routes in *Thermococcus kodakarensis* linking glyceraldehyde 3-phosphate and 3-phosphoglycerate. Mol Microbiol **81:**1300-12.

42. **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.** 2009. Novel multiprotein complexes identified in the hyperthermophilic archaeon *Pyrococcus furiosus* by non-denaturing fractionation of the native proteome. Mol Cell Proteomics **8:**735-51.

43. **Mukund, S., and M. W. Adams.** 1995. Glyceraldehyde-3-phosphate ferredoxin oxidoreductase, a novel tungsten-containing enzyme with a potential glycolytic role in the hyperthermophilic archaeon *Pyrococcus furiosus*. J Biol Chem **270:**8389-92.

44. **Namba, K., K. Hagiwara, H. Tanaka, Y. Nakaishi, K. T. Chong, E. Yamashita, G. E. Armah, Y. Ono, Y. Ishino, T. Omura, T. Tsukihara, and A. Nakagawa.** 2005. Expression and molecular characterization of spherical particles derived from the genome of the hyperthermophilic euryarchaeote *Pyrococcus furiosus*. J Biochem **138:**193-9.

45. **Noll, K. M., P. Lapierre, J. P. Gogarten, and D. M. Nanavati.** 2008. Evolution of mal ABC transporter operons in the *Thermococcales* and *Thermotogales*. BMC Evol Biol **8:**7.

46. **Peak, M. J., F. T. Robb, and J. G. Peak.** 1995. Extreme resistance to thermally induced DNA backbone breaks in the hyperthermophilic archaeon *Pyrococcus furiosus*. J Bacteriol **177:**6316-8.

47. **Reher, M., S. Gebhard, and P. Schonheit.** 2007. Glyceraldehyde-3-phosphate ferredoxin oxidoreductase (GAPOR) and nonphosphorylating glyceraldehyde-3-phosphate dehydrogenase (GAPN), key enzymes of the respective modified Embden-Meyerhof pathways in the hyperthermophilic crenarchaeota *Pyrobaculum aerophilum* and *Aeropyrum pernix*. FEMS Microbiol Lett **273:**196-205.

48. **Rice, P., I. Longden, and A. Bleasby.** 2000. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet **16:**276-7.

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

50. **Santangelo, T. J., L. Cubonova, and J. N. Reeve.** 2010. *Thermococcus kodakarensis* genetics: TK1827-encoded beta-glycosidase, new positive-selection protocol, and targeted and repetitive deletion technology. Appl Environ Microbiol **76:**1044-52.

51. **Schut, G. J., S. D. Brehm, S. Datta, and M. W. Adams.** 2003. Whole-genome DNA microarray analysis of a hyperthermophile and an archaeon: *Pyrococcus furiosus* grown on carbohydrates or peptides. J Bacteriol **185:**3935-47.

52. **Siguier, P., J. Perochon, L. Lestrade, J. Mahillon, and M. Chandler.** 2006. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res **34:**D32-6.

53. **Srivatsan, A., Y. Han, J. Peng, A. K. Tehranchi, R. Gibbs, J. D. Wang, and R. Chen.** 2008. High-precision, whole-genome sequencing of laboratory strains facilitates genetic studies. PLoS Genet **4:**e1000139.

54. **Strand, K. R., C. Sun, T. Li, F. E. Jenney, Jr., G. J. Schut, and M. W. Adams.** 2010. Oxidative stress protection and the repair response to hydrogen peroxide in the hyperthermophilic archaeon *Pyrococcus furiosus* and in related species. Arch Microbiol **192:**447-59.

55. **Tran, T. T., P. Dam, Z. Su, F. L. Poole, 2nd, M. W. Adams, G. T. Zhou, and Y. Xu.** 2007. Operon prediction in *Pyrococcus furiosus*. Nucleic Acids Res **35:**11-20.

56. **Ward, D. E., W. M. de Vos, and J. van der Oost.** 2002. Molecular analysis of the role of two aromatic aminotransferases and a broad-specificity aspartate aminotransferase in the aromatic amino acid metabolism of *Pyrococcus furiosus*. Archaea **1:**133-41.

57. **Weinberg, M. V., G. J. Schut, S. Brehm, S. Datta, and M. W. Adams.** 2005. 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 **187:**336-48.

58. **White, J. R., P. Escobar-Paramo, E. F. Mongodin, K. E. Nelson, and J. DiRuggiero.** 2008. Extensive genome rearrangements and multiple horizontal gene transfers in a population of *Pyrococcus* isolates from Vulcano Island, Italy. Appl Environ Microbiol **74:**6447-51.

59. **Yoon, S. H., D. J. Reiss, J. C. Bare, D. Tenenbaum, M. Pan, J. Slagel, R. L. Moritz, S. Lim, M. Hackett, A. L. Menon, M. W. Adams, A. Barnebey, S. M. Yannone, J. A. Leigh, and N. S. Baliga.** 2011. Parallel evolution of transcriptome architecture during genome reorganization. Genome Res **21:**1892-904.

60. **Zivanovic, Y., P. Lopez, H. Philippe, and P. Forterre.** 2002. *Pyrococcus* genome comparison evidences chromosome shuffling-driven evolution. Nucleic Acids Res **30:**1902-10.

Table A.1. Comparison of IS element population in *P. furiosus* COM1 strain relative to the NCBI reference sequence

IS Family	Name	COM1 Genome		NCBI Reference	
		Full	Partial	Full	Partial
IS6	ISPfu1	16	1	8	1
	ISPfu2	13	0	11	0
	ISPfu5	4	0	4	0
IS982	ISPfu3	5	0	5	0
IS607	ISPfu4	1	2	1	2
	ISTko2	0	2	0	2
IS200/IS605	ISTsi3	0	1	0	1
	Total IS elements	39	6	29	6
	Total IS lengths (bp)	32406	1798	24592	1798
	Genome (%)	1.70	0.09	1.29	0.09

ISFinder blast analysis tool (52) was used to annotate IS elements with e-value scores $\leq$ 4E-17 in both the COM1 and NCBI reference genome sequences. Sequence similarity among the members of an IS family and between IS families results in multiple hits at each locus. Only the best BLAST hit was retained for each IS element.

Table A.2. Genes affected by IS activity in the *P. furiosus* COM1 strain

IS Element[a]	Affected Gene(s)	Gene Annotation[b]	Length (bp)	Operon Position[c]	Predicted Location[d]
Excisions[e]			**Deletion**		
PF0013 (ISPfu1)	PF0012	hypothetical protein	39 (3' end)		cyt
PF0408 (ISPfu1)	PF0407	<carboxypeptidase-like, regulatory domain>	41 (5' end)	1 of 2	mem
PF0756 (ISPfu2)	PF0755	non-phosphorylating GAPDH (GAPN)	300 (5'end)	4 of 4	cyt
	PF0757	hypothetical protein	complete		cyt
	PF0758	<nucleotide-binding domain (HEPN)>	complete		cyt
PF2035 (ISPfu2)	PF2034	<methyltransferase (TYW3)>	41 (5'end)	1 of 2	cyt
Insertions in ORF[f]			**IS Element**		
ISPfu2	PF0061	riboflavin synthase subunit alpha	782	1 of 4	cyt
ISPfu1	PF0393	[CRISPR-associated protein, Cas6-2 (21)]	781		cyt
ISPfu1	PF0429	proline permease	781		mem
ISPfu1	PF0823	<multi antimicrobial extrusion protein (MatE)>	781	1 of 2	mem
ISPfu1	PF1260	hypothetical protein	781	1 of 2	cyt
ISPfu1	PF1603	Na antiporter	781	1 of 3	mem
ISPfu1	PF2059	aminopeptidase	781	1 of 2	exc
Insertion in Intergenic Regions[g]			**IS Element**		
ISPfu1	PF0189	dihydroorotase	781	3 of 3	cyt
	PF0190	[cold-induced protein A, CipA (57)]			mem

ISPfu1	PF0271	hypothetical protein	781		exc
	PF0272	[4-α-glucanotransferase, (35)]			cyt
ISPfu1	PF0401	methyltransferase	781	4 of 4	exc
ISPfu1	PF1738	sugar kinase	781		cyt
	PF1739	[Mal I transporter, (31)]		1 of 6	exc
ISPfu2	PF_t006	tRNA Glu anticodon TTC	782		
ISPfu2	PF0497	<Winged helix-turn-helix transcription factor>	782	1 of 9	cyt
ISPfu2	PF1239	hypothetical protein	782		mem
	PF1240	purine permease		1 of 2	mem

[a] IS element nomenclature based on ISFinder database (52).

[b] Annotations are NCBI gene names (no brackets) and literature cited ([ ]), except for hypothetical genes, where the best IPR (< >) match is given if available.

[c] Operon predictions from (55) based on NCBI reference sequence.

[d] Predicted cellular location based on predicted TMDs, proteins with ≥ 2 TMDs classified as membrane (mem), proteins with <2 TMDs and a predicted signal peptide using SignalP with the gram-positive model ≥ 0.6 are classified as extracellular (exc), and proteins with <2 TMDs and no predicted signal peptide are classified as cytoplasmic (cyt).

[e] Includes neighboring genes affected by IS excision, either full deletion (complete) or truncation (bp) at the chromosomal-level.

[f] Includes genes disrupted by IS insertion within their open reading frame.

[g] Includes neighboring genes with transcriptional start sites downstream of IS insertion.

Table A.3. Additional large chromosomal deletions in *P. furiosus* COM1 strain relative to NCBI reference

Gene	Gene Annotation[b]	Deletion Length (bp)	Operon Position[c]	Predicted Location[d]
PF0073	beta-glucosidase (CelB)	774 of 1,419 (3' end)		cyt
PF1249[h]	ABC transporter	complete	1 of 2	cyt
PF1250	hypothetical protein	complete	2 of 2	exc
PF1251	<alpha-galactosidase, NEW3 domain>	complete	1 of 2	cyt
PF1252	hypothetical protein	complete	2 of 2	mem
PF1253	[aromatic aminotransferase II (56)]	complete		cyt
PF1254	sodium dependent transporter	904 of 1,737 (5' end)	1 of 2	mem

[b] Annotations are NCBI gene names (no brackets) and literature cited ([ ]), except for hypothetical genes, where the best IPR (< >) match is given if available.

[c] Operon predictions from (55) based on NCBI reference sequence.

[d] Predicted cellular location, for explanation see Table 2, footnote d.

[h] Adjacent to transposase (PF1248, ISPfu3) in both COM1 and reference sequence.

Table A.4. Major protein-level genome differences in P. furiosus COM1 strain relative to NCBI reference (<90% identity)

Gene	Gene Annotation[b]	Identical Residues (AA)[i]	Alignment Length (AA)[i]	Identity (%)[i]	Operon Position[c]	Predicted Location[d]
PF0054	AsnC family transcriptional regulator	91	155	58.7	6 of 7	cyt
PF0067	cobalt ABC transporter	157	243	64.6	1 of 2	mem
PF0147	potassium channel-like protein	23	205	11.2	2 of 2	cyt
PF0225	N-acetyltransferase	111	201	55.2	1 of 2	cyt
PF0228	hypothetical protein	49	89	55.1	2 of 3	cyt
PF0244	hypothetical protein	38	89	42.7		cyt
PF0314	signal peptidase	58	173	33.5	2 of 3	cyt
PF0334	flagella-like protein	24	194	12.4	2 of 6	exc
PF0351	hypothetical protein	22	285	7.7		cyt
PF0352	[CRISPR-associated protein, Cmr1-2 (21)]	189	451	41.9		cyt
PF0363	beta-galactosidase	11	775	1.4	7 of 7	cyt
PF0423	hypothetical protein	45	117	38.5	1 of 3	cyt
PF0424	hypothetical protein	34	172	19.8	2 of 3	cyt
PF0439	hypothetical protein	297	401	74.1		cyt
PF0472	<class III signal peptide motif>	42	54	77.8	2 of 2	exc
PF0524	<ribbon-helix-helix (Met_repress_like)>	57	72	79.2	1 of 4	cyt
PF0611	hypothetical protein	295	382	77.2		cyt

PF0622	hypothetical protein	7	33	21.2		cyt
PF0710	<transcription repressor DNA-binding>	79	157	50.3	3 of 3	cyt
PF0764	DEXX-box atpase	232	434	53.5		cyt
PF0777	hypothetical protein	191	215	88.8		mem
PF0785.2n	hypothetical protein	148	192	77.1		N/A
PF0873	hypothetical protein	36	196	18.4		cyt
PF0874	membrane dipeptidase	9	379	2.4	1 of 2	cyt
PF0901	hypothetical protein	121	142	85.2		cyt
PF0960	hypothetical protein	15	112	13.4	2 of 2	exc
PF1075	hypothetical protein	3	219	1.4		mem
PF1109	<carbohydrate-binding domain (CARDB)>	959	1132	84.7	1 of 2	exc
PF1113	<tRNA amidotransferease GAD domain>	199	223	89.2	1 of 3	cyt
PF1182	<thioredoxin-like fold>	7	126	5.6		cyt
PF1182.1n	hypothetical protein	53	169	31.4		N/A
PF1206	<nucleic acid-binding, PIN, PH0500>	15	156	9.6	1 of 3	cyt
PF1748	sulfate permease, ABC transporter	321	543	59.1	4 of 6	mem
PF1749	sulfate transport integral membrane protein	10	228	4.4	5 of 6	mem
PF1761	<putative cell wall binding repeat 2>	204	389	52.4	1 of 2	exc
PF1935[j]	Amylopullulanase	1114	1355	82.2	3 of 6	exc

[b] Annotations are NCBI gene names (no brackets) and literature cited ([ ]), except for hypothetical genes, where the best IPR (< >) match is given if available.

[c] Operon predictions from (55) based on NCBI reference sequence.

[d] Predicted cellular location, for explanation see Table 2, footnote d.

[i] Based on Needleman-Wunsch global alignment with NCBI reference sequence.

[j] A fusion of PF1934 and PF1935 was previously reported (35). A longer product was found in COM1 which is closer in length to those in other *Pyrococci*.

Figure A.1. Dot plot of the alignment of the *P. furiosus* NCBI reference sequence (NC_003413) on the x-axis and the *P. furiosus* COM1 strain on the y-axis indicates high overall synteny with two major inversions. Inverted segments in COM1 synonymous with the circular diagram in Fig. 2 are: block A (1,497,249-1,297,198), block E (1,911,081-1,497,250), and block C (199,582-162,245). The small segment (210,862-211,700) is a result of an IS element insertion between PF0401 and PF0402 in COM1.

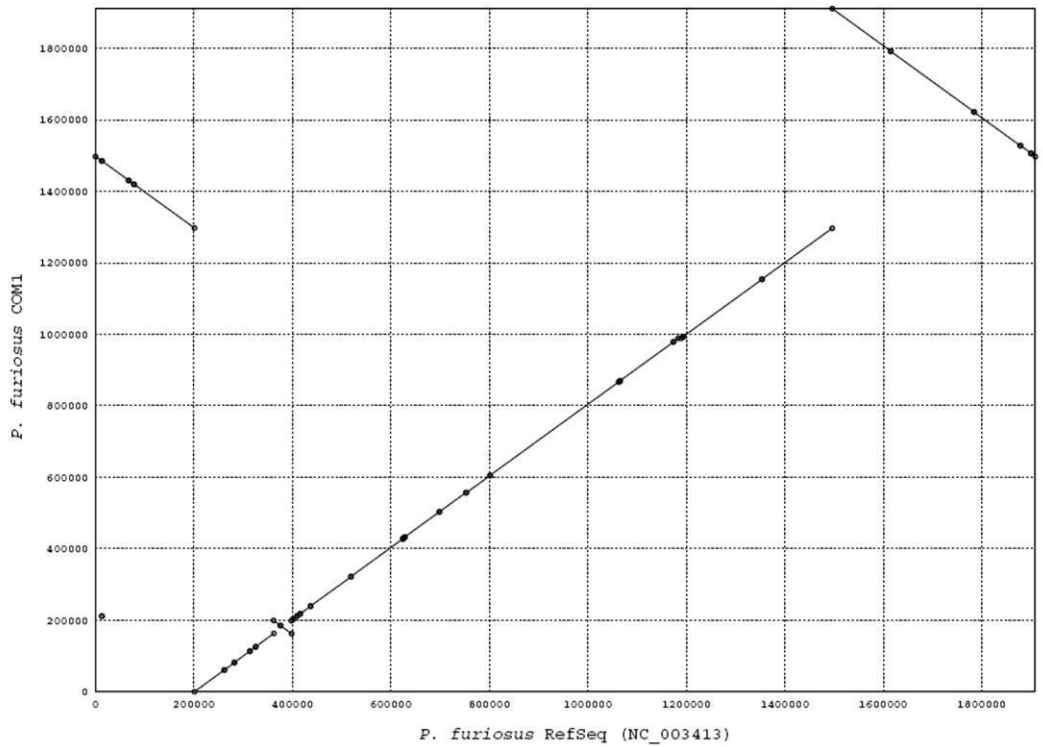
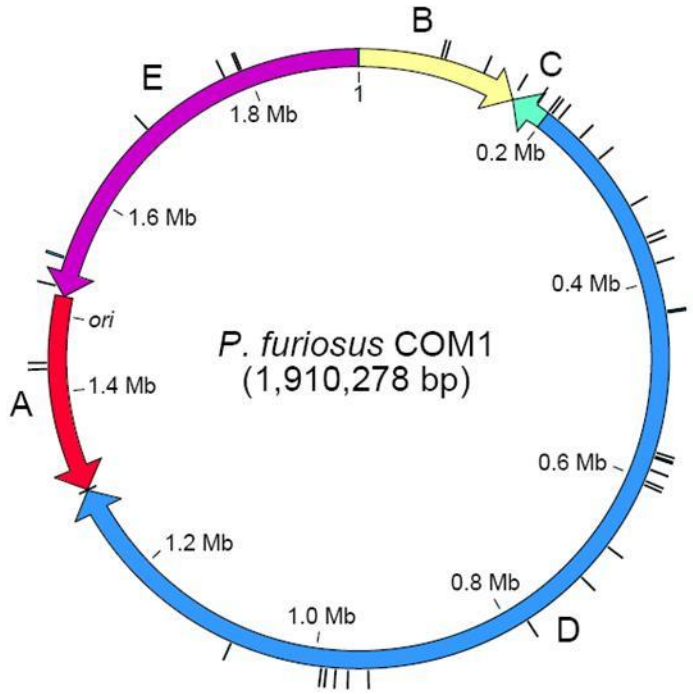

P. furiosus COM1
P. furiosus RefSeq (NC_003413)
0
200000
400000
600000
800000
1000000
1200000
1400000
1600000
1800000

Figure A.2. COM1 genome organization (panel A) compared to the *P. furiosus* reference sequence (NC_003413, panel B). Black tick marks on the outside of the circular diagrams represent full length IS elements in both genomes, with two additional IS elements located at the boundaries on the COM1 genome between blocks A and D, and B and E. The block boundary between A and E indicates start/end of reference sequence and is just upstream of the origin of replication (ori) in both genomes. PF numbers associated with each block are ordered in the 5' to 3' direction of each arrow: A (red) PF0001 – PF0189, B (yellow) PF0190 – PF0347, C (green) PF0349 – PF0388, D (blue) PF0390 – PF1603', E (purple) PF1603'' – PF2065. PF1603 is disrupted by an IS element in the COM1 genome.

A


B
C
E
1
1.8 Mb
0.2 Mb
1.6 Mb
0.4 Mb
ori
P. furiosus COM1
(1,910,278 bp)
A
1.4 Mb
0.6 Mb
1.2 Mb
0.8 Mb
1.0 Mb
D

B

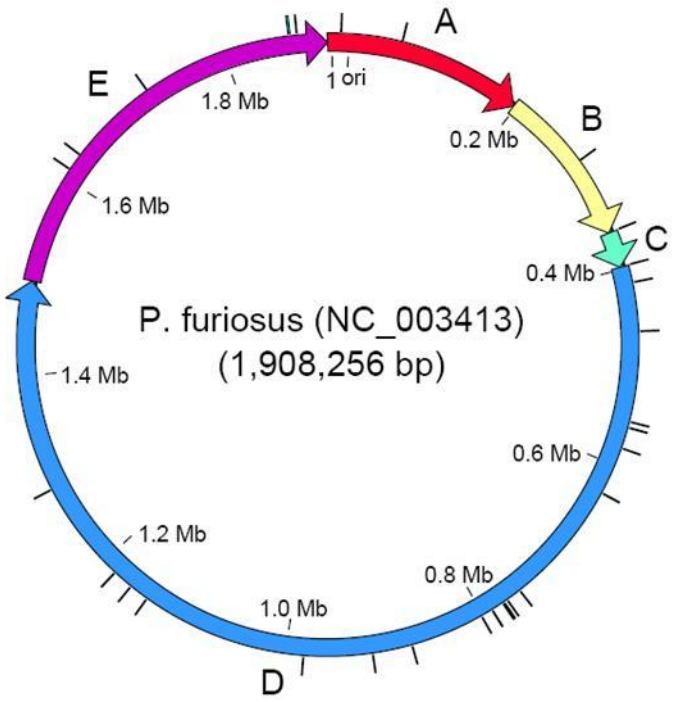
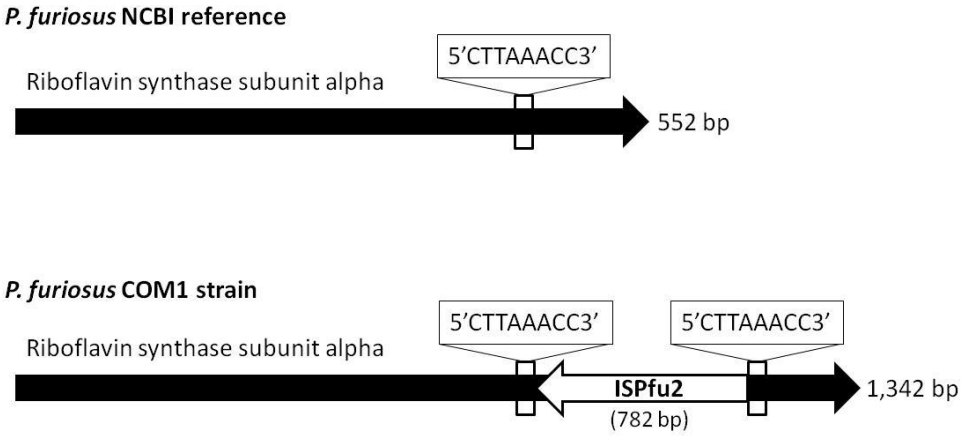

A
B
C
D
E
1 ori
0.2 Mb
0.4 Mb
0.6 Mb
0.8 Mb
1.0 Mb
1.2 Mb
1.4 Mb
1.6 Mb
1.8 Mb
P. furiosus (NC_003413)
(1,908,256 bp)

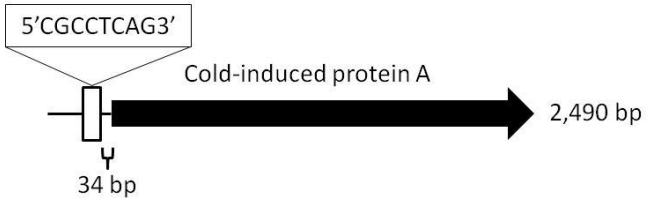
Figure A.3. Maps of IS element insertions in both coding and potential regulatory regions of the COM1 genome compared to NCBI reference sequence. (A) the riboflavin synthase subunit alpha (PF0061) gene and (B) the region upstream of the cold-induced protein A (PF0190) in COM1 are disrupted by IS elements. IS elements are represented by white arrows and are flanked by direct repeat sequences (boxed). A single copy of the direct repeat sequence is located at the same positions on the reference sequences and the nucleotide sequence alignments are identical for both genomic regions minus the IS element insertion and additional direct repeats. In the case of the riboflavin synthase gene,at the protein-level the COM1 sequence has premature stop codon (~16 bp into the IS element).

A

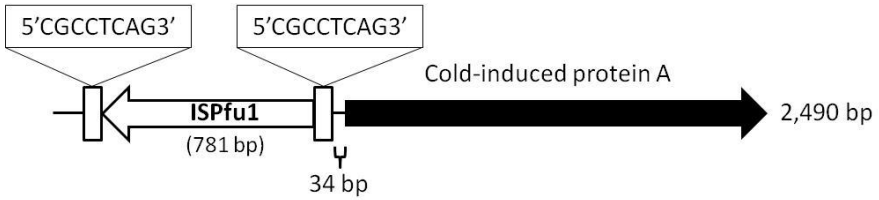

P. furiosus NCBI reference
Riboflavin synthase subunit alpha
5'CTTAAACC3'
552 bp
P. furiosus COM1 strain
Riboflavin synthase subunit alpha
5'CTTAAACC3'
5'CTTAAACC3'
ISPfu2
(782 bp)
1,342 bp

B

***P. furiosus* NCBI reference**



***P. furiosus* COM1 strain**



SUPPORTING MATERIAL

Table A.S1. Disrupted genes in *P. furiosus* COM1 strain relative to NCBI reference. These include genes IS-mediated and IS-independent deletions, major chromosome-level deletions, major (<90% identity) and minor (≥90% identity) protein-level genome differences, and silent mutations.

Gene	Gene Annotation[b]	Identical Residues (AA)[i]	Alignment Length (AA)[i]	Identity (%)[i]	Operon Position[c]	Location[d]
PF0002	<YcaO-like>	433	434	99.8	2 of 5	cyt
PF0012[k]	hypothetical protein	252	267	94.4		cyt
PF0036	daunorubicin resistance membrane protein	262	263	99.6	2 of 2	mem
PF0054[m]	AsnC family transcriptional regulator	91	155	58.7	6 of 7	cyt
PF0057	<peptidase M28>	538	551	97.6		cyt
PF0058	dolichol monophosphate mannose synthase	351	352	99.7	1 of 3	mem
PF0061[k]	riboflavin synthase subunit alpha	163	186	87.6	1 of 4	cyt
PF0067[m]	cobalt transport ABC transporter	157	243	64.6	1 of 2	mem
PF0073[l]	beta-glucosidase (CelB)	209	472	44.3		cyt
PF0079	cyclic 2,3-diphospoglycerate-synthetase	430	431	99.8	2 of 2	cyt
PF0085	DNA helicase	1274	1355	94	1 of 2	cyt
PF0122	translation initiation factor IF-2	320	321	99.7	1 of 2	cyt
PF0132	hypothetical protein	486	489	99.4	1 of 2	cyt
PF0147[m]	potassium channel-like protein	23	205	11.2	2 of 2	cyt
PF0179	V-type ATP synthase subunit E	202	203	99.5	7 of 13	cyt
PF0209	ribosomal protein s6 modification protein	271	273	99.3	3 of 3	cyt

PF0220	hexulose-6-phosphate synthase	433	454	95.4		cyt
PF0225[m]	N-acetyltransferase	111	201	55.2	1 of 2	cyt
PF0228[m]	hypothetical protein	49	89	55.1	2 of 3	cyt
PF0244[m]	hypothetical protein	38	89	42.7		cyt
PF0270	alanyl-tRNA synthetase	912	914	99.8		cyt
PF0314[m]	signal peptidase	58	173	33.5	2 of 3	cyt
PF0332	flagellar accessory protein FlaH	203	204	99.5	4 of 4	cyt
PF0334[m]	flagella-like protein	24	194	12.4	2 of 6	exc
PF0351[m]	hypothetical protein	22	285	7.7		cyt
PF0352[m]	[CRISPR-associated protein, Cmr1-2 (21)]	189	451	41.9		cyt
PF0356	beta-galactosidase	481	483	99.6	3 of 3	cyt
PF0357	dipeptide-binding protein	630	631	99.8	1 of 7	mem
PF0360	oligopeptide ABC transporter	323	324	99.7	4 of 7	cyt
PF0363[m]	beta-galactosidase	11	775	1.4	7 of 7	cyt
PF0393[k]	[CRISPR-associated protein, Cas6-2 (21)]	11	241	4.6		cyt
PF0404[n]	<von Willebrand factor, type A (VWA)>	418	418	100	3 of 5	exc
PF0407[k]	<carboxypeptidase-like, regulatory domain>	14	607	2.3	1 of 2	mem
PF0423[m]	hypothetical protein	45	117	38.5	1 of 3	cyt
PF0424[m]	hypothetical protein	34	172	19.8	2 of 3	cyt
PF0429[k]	proline permease	195	493	39.6		mem
PF0439[m]	hypothetical protein	297	401	74.1		cyt
PF0445[n]	galactokinase	352	352	100	3 of 4	cyt
PF0446	NDP-sugar synthase	151	152	99.3	4 of 4	cyt

PF0463	hydrolase related to 2-haloalkanoic acid dehalogenase	233	234	99.6	1 of 5	cyt
PF0472 [m]	<class III signal peptide motif>	42	54	77.8	2 of 2	exc
PF0509	integral membrane glycosyltransferase	673	707	95.2	2 of 4	mem
PF0524 [m]	<ribbon-helix-helix (Met_repress_like)>	57	72	79.2	1 of 4	cyt
PF0534	indolepyruvate oxidoreductase beta	197	214	92.1	3 of 3	cyt
PF0552	arsenical-resistance protein acr3	363	399	91		mem
PF0572	dna2-nam7 helicase family protein	654	655	99.8		cyt
PF0592	ATP-dependent RNA helicase	866	867	99.9	5 of 6	cyt
PF0600	hypothetical protein	354	356	99.4	1 of 3	exc
PF0611 [m]	hypothetical protein	295	382	77.2		cyt
PF0621	<winged helix-turn-helix transcription repressor DNA-binding>	255	283	90.1	2 of 2	exc
PF0622 [m]	hypothetical protein	7	33	21.2		cyt
PF0710 [m]	<transcription repressor DNA-binding>	79	157	50.3	3 of 3	cyt
PF0714	<flavodoxin/nitric oxide synthase>	169	176	96		cyt
PF0755 [k]	non-phosphorylating GAPDH (GAPN)	389	506	76.9	4 of 4	cyt
PF0757 [k]	hypothetical protein	0	0	0		cyt
PF0758 [k]	<nucleotide-binding domain (HEPN)>	0	0	0		cyt
PF0764 [m]	DEXX-box atpase	232	434	53.5		cyt
PF0777 [m]	hypothetical protein	191	215	88.8		mem
PF0782	<polysaccharide biosynthesis protein>	107	109	98.2	1 of 2	mem
PF0785.2n	hypothetical protein	148	192	77.1		N/A

[m]

PF0793	hypothetical protein	81	87	93.1	7 of 7	cyt
PF0823[k]	<multi antimicrobial extrusion protein (MatE)>	115	472	24.4	1 of 2	mem
PF0856	isopentenyl pyrophosphate isomerase	370	394	93.9		cyt
PF0868[n]	NDP-sugar synthase	413	413	100		cyt
PF0871[n]	hypothetical protein	135	135	100	2 of 3	cyt
PF0872	<circadian clock protein KaiC/DNA repair protein RadA>	115	125	92	3 of 3	cyt
PF0873[m]	hypothetical protein	36	196	18.4		cyt
PF0874[m]	membrane dipeptidase	9	379	2.4	1 of 2	cyt
PF0884	hypothetical protein	346	347	99.7	4 of 4	cyt
PF0889[n]	hypothetical protein	325	325	100		exc
PF0901[m]	hypothetical protein	121	142	85.2		cyt
PF0921	ABC transporter	285	305	93.4	1 of 5	cyt
PF0933	DNA repair helicase Rad3	626	647	96.8	3 of 3	cyt
PF0960[m]	hypothetical protein	15	112	13.4	2 of 2	exc
PF0978	<transglutaminase-like domain>	314	328	95.7	7 of 7	exc
PF1030	methionyl-tRNA synthetase	723	724	99.9		cyt
PF1075[m]	hypothetical protein	3	219	1.4		mem
PF1109[m]	<carbohydrate-binding domain (CARDB)>	959	1132	84.7	1 of 2	exc
PF1113[m]	<tRNA amidotransferease GAD domain>	199	223	89.2	1 of 3	cyt
PF1114[o]	orotidine 5'-phosphate decarboxylase	0	0	0	2 of 3	cyt
PF1120	ATP-dependent RNA helicase	723	724	99.9	3 of 4	cyt
PF1121	[CRISPR-associated protein, Cas5t (21)]	229	230	99.6	4 of 4	cyt

PF1130	[CRISPR-associated protein, Cmr1-1 (21)]	325	338	96.2	7 of 7	cyt
PF1131	[CRISPR-associated protein, Cas6 (21)]	263	264	99.6		cyt
PF1159	L-tyrosine decarboxylase	362	387	93.5	2 of 4	cyt
PF1168	[5’ to 3’ exonuclease, NurA (23)]	450	451	99.8	6 of 7	exc
PF1182[m]	<thioredoxin-like fold>	7	126	5.6		cyt
PF1182.1n[m]	hypothetical protein	53	169	31.4		N/A
PF1206[m]	<nucleic acid-binding, PIN, PH0500>	15	156	9.6	1 of 3	cyt
PF1208	beta-mannosidase	509	510	99.8	3 of 3	cyt
PF1225	hypothetical protein	94	95	98.9	2 of 3	cyt
PF1238	ABC transporter	622	632	98.4	2 of 2	cyt
PF1247.1n	hypothetical protein	151	152	99.3		N/A
PF1249[l]	ABC transporter	0	0	0	1 of 2	cyt
PF1250[l]	hypothetical protein	0	0	0	2 of 2	exc
PF1251[l]	<alpha-galactosidase, NEW3 domain>	0	0	0	1 of 2	cyt
PF1252[l]	hypothetical protein	0	0	0	2 of 2	mem
PF1253[l]	[aromatic aminotransferase II (56)]	0	0	0		cyt
PF1254[l]	sodium dependent transporter	270	578	46.7	1 of 2	mem
PF1258	ribose-5-phosphate isomerase A	222	229	96.9	2 of 3	cyt
PF1259	hypothetical protein	106	107	99.1	3 of 3	mem
PF1260[k]	hypothetical protein	183	304	60.2	1 of 2	cyt
PF1268[n]	5-methyltetrahydropteroyltriglutamate--homocysteine S-methyltransferase	309	309	100	3 of 4	cyt
PF1305	hypothetical protein	717	721	99.4	4 of 5	exc

PF1328	ferredoxin-NADP reductase	277	278	99.6	3 of 3	cyt
PF1407	uridylate kinase	224	225	99.6	4 of 4	cyt
PF1433[n]	membrane bound hydrogenase beta	173	173	100	11 of 14	cyt
PF1434[n]	membrane bound hydrogenase alpha	427	427	100	12 of 14	cyt
PF1480	formaldehyde:ferredoxin oxidoreductase wor5	570	633	90	2 of 2	cyt
PF1494	hypothetical protein	190	193	98.4		mem
PF1567	exosome complex RNA-binding protein Rrp42	276	277	99.6	1 of 5	cyt
PF1572	<amino acid-binding ACT (ACT)>	174	175	99.4		cyt
PF1573[n]	<peptidase U62, modulator of DNA gyrase>	455	455	100	1 of 2	cyt
PF1589	<carbohydrate/purine kinase (PfkB)>	250	251	99.6		cyt
PF1603[k]	Na antiporter	185	427	43.3	1 of 3	mem
PF1615	<intein DOD homing endonuclease>	969	970	99.9	4 of 4	cyt
PF1624	hypothetical protein	174	175	99.4		mem
PF1718	tRNA-modifying protein	337	338	99.7		cyt
PF1725	DNA primase	422	452	93.4	2 of 2	cyt
PF1748[m]	system permease, ABC transporter	321	543	59.1	4 of 6	mem
PF1749[m]	sulfate transport membrane protein	10	228	4.4	5 of 6	mem
PF1761[m]	<putative cell wall binding repeat 2>	204	389	52.4	1 of 2	exc
PF1774	iron ABC transporter ATP-binding protein	342	363	94.2		exc
PF1882	cell division protein CDC48	795	796	99.9		cyt
PF1896.1n	hypothetical protein	63	64	98.4		N/A
PF1902	DNA repair helicase	444	447	99.3		cyt
PF1910[n]	oxidoreductase	476	476	100	1 of 2	cyt

PF1935[j m]	amylopullulanase	1114	1355	82.2	3 of 6	exc
PF1971	anaerobic ribonucleoside triphosphate reductase	604	605	99.8	1 of 2	cyt
PF1989	preprotein translocase subunit SecE	60	61	98.4	1 of 6	cyt
PF2000	glycine dehydrogenase subunit 2	494	502	98.4	2 of 2	cyt
PF2034[k]	<methyltransferase (TYW3)>	206	229	90	1 of 2	cyt
PF2059[k]	aminopeptidase	29	567	5.1	1 of 2	exc

[b] Annotations are NCBI gene names (no brackets) and literature cited ([ ]), except for hypothetical genes, where the best IPR (< >) match is given if available.
[c] Operon predictions from (55) based on NCBI reference sequence.
[d] Predicted cellular location based on predicted TMDs, proteins with >= 2 TMDs classified as membrane (mem), proteins with <2 TMDs and a predicted signal peptide using SignalP with the gram-positive model >= 0.6 are classified as extracellular (exc), and proteins with <2 TMDs and no predicted signal peptide are classified as cytoplasmic (cyt).
[i] Based on Needleman-Wunsch global alignment with NCBI reference sequence.
[j] A fusion of PF1934 and PF1935 was previously reported (35). A longer product was found in COM1 which is closer in length to those in other *Pyrococci*.
[k] Genes affected by IS activity (Table 2).
[l] Additional large chromosomal deletions (Table 3).
[m] Major protein-level genome differences (Table 4).
[n] Silent mutation at protein-level.
[o] Auxotrophic strain deletion (37).

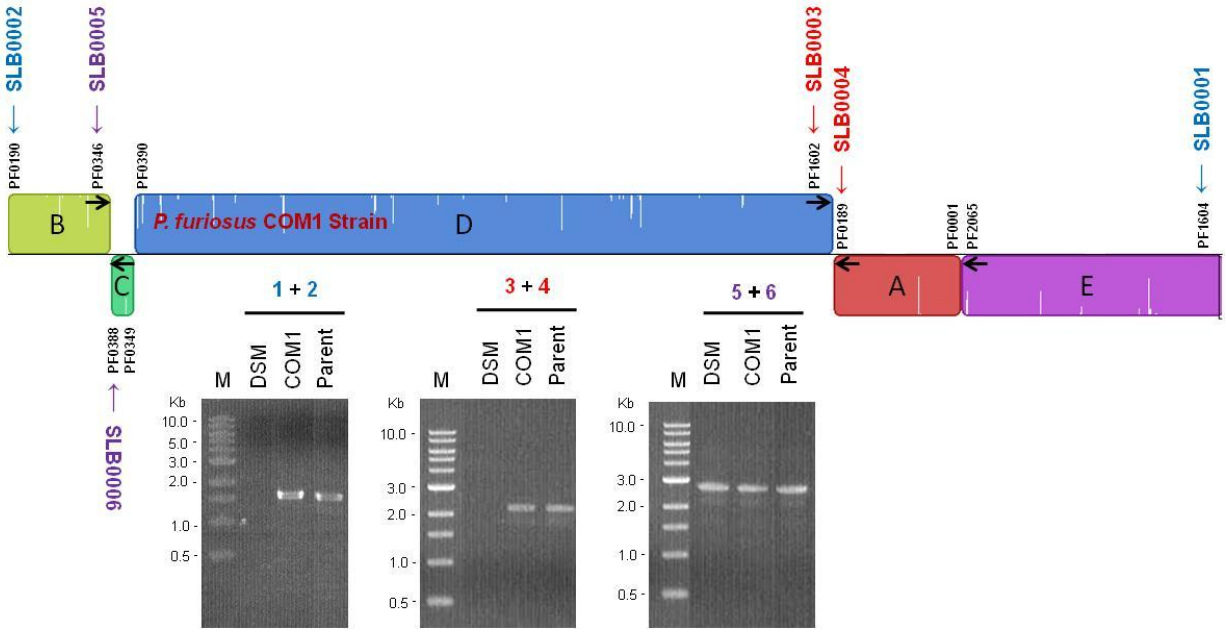


Figure A.S1. PCR confirmation of COM1 genome organization. PCR primers (SLB0001-6) were designed within ORFs up or downstream of block boundaries. PCR products were sequenced and further confirmed COM1 genome rearrangement. Inverted orientations of block C and block AE was also observed in the Parent. Only the inverted order of block C was observed in the DSMZ strain. Primer sequences are as follows: SLB0001 (5'- AGG CTA ATA AGG GTA GGT GG -3‘), SLB0002 (5'- AAG TGG GCT TAC TAA CAT GC -3‘), SLB0003 (5'- ATC TCC ATA TAT TGG GCA GC -3‘), SLB0004 (5'- ACA ACT TCC ATG CTA ATC ACC -3‘), SLB0005 (5'- TCT TAG CCA ATT CTT CGT CG -3’), and SLB0006 (5'- TCT AAC AGG AGT CAC TGC C -3’).

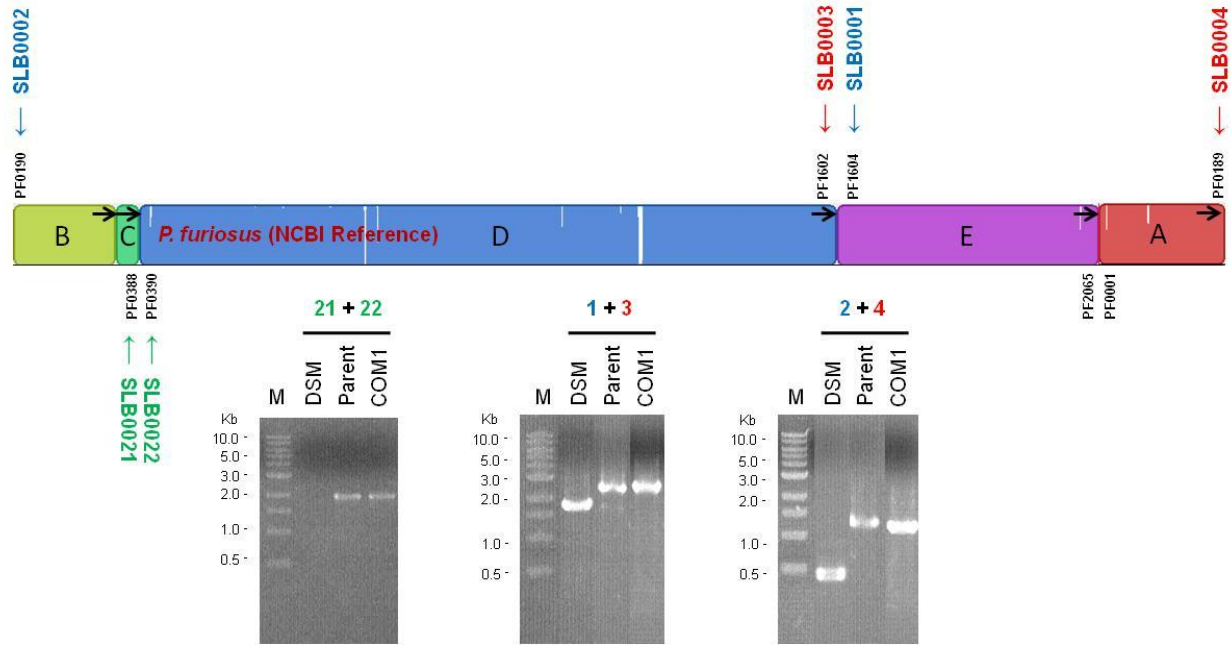


Figure A.S2. PCR analysis of NCBI reference genome organization. PCR primers (SLB0001-4 and 21-22) were designed within ORFs up or downstream of block boundaries. PCR products were sequenced to confirm genome arrangements. NCBI reference arrangement of blocks AE and C were observed in the COM1 and Parent with IS element insertion in amplified products. Only the inverted order of block C (Figure S1) and the reference order of block AE were observed in the DSMZ strain. Primer sequences are as follows: SLB0001 (5'- AGG CTA ATA AGG GTA GGT GG -3‘), SLB0002 (5'- AAG TGG GCT TAC TAA CAT GC -3‘), SLB0003 (5'- ATC TCC ATA TAT TGG GCA GC -3‘), SLB0004 (5'- ACA ACT TCC ATG CTA ATC ACC -3‘), SLB0021 (5'- TCA GTA TTG CGA TGA GAG C -3’), and SLB0022 (5'- TCC ATT CTC GAC AAT CTC TCC -3’).

APPENDIX B

METALLOMICS USING INDUCTIVELY COUPLED PLASMA MASS SPECTROMETRY[5]

[5] Vaccaro BJ, Menon AL, Lancaster WA, Adams, MWW. *Accepted, in press for Current Protocols in Chemical Biology*.

ABSTRACT

Inductively Coupled Plasma Mass Spectrometry (ICP-MS) is a highly sensitive elemental analysis technique that has been widely applied in many fields. Here we describe applications using a broad-scale approach to examine metal usage in biology. These protocols address questions such as: Which elements from the surrounding environment are taken up into the cells of a given organism? How does this vary between different organisms? Which metals are "bound" and which are "free"? What sort of proteins are the "bound" metals associated with? This allows for investigations into several branches of bioinorganic chemistry including uptake, toxicity and detoxification, bioremediation and the discovery of new uses of elements. In the protocols that are presented, there is an emphasis on metals and, more narrowly, on transition metals because these comprise the majority of tightly protein-bound, low abundance elements. Non-metals, metalloids, main-group metals, and f-block metals were also analyzed and investigated. The sample preparation procedure requires acid-lability for detection, which likely eliminates certain nonmetals, such as selenium. However, one advantage of the protocols described is that they are quite readily adapted to measure any element of interest.

Keywords: metalloproteome, metallome, ICP-MS, metalloprotein, metalloenzyme, bioinorganic chemistry, metallobiochemistry

INTRODUCTION

Low abundance elements play a crucial yet somewhat poorly understood role in biology. There are several elements that are known to be essential to a wide range of organisms, but their

role is largely or wholly unknown. These include chromium (Cr), vanadium (V), boron (B), silicon (Si), fluorine (F), and iodine (I) (2). Indeed, even for humans, the list of metals that are absolutely required is still not certain. The use of these elements, quite naturally, takes advantage of their unique chemical and physical properties, and they often play extraordinary roles in their biological context. For instance, metalloenzymes perform the most industrially-relevant, the most energetically challenging and arguably the most interesting reactions in all biology. Nitrogen fixation (catalyzed by nitrogenase) (17), the interconversion of CO and CO_2 (carbon monoxide dehydrogenase) (16), photolysis of water (oxygen-evolving complex) (9), hydrogen production (hydrogenase) (13), C-H activation (methane monooxygenase and cytochrome P450s) (4), radical-based chemistry (radical SAM enzymes and cobalamin) (7) (14) and electron transport (15) are all carried out biologically using metal centers. These are reactions with applications in agriculture, environmental remediation, biofuels, and synthetic chemistry. Moreover, they have had a major influence on our understanding of chemistry in general.

A complete perspective of bioinorganic chemistry involves much more than the constructive use of elements in biological processes, and the protocols presented here are equally applicable to the study of detrimental effects or toxicity of metals. Indeed, elements are often classified as either "essential" or "toxic," but this is entirely too simplistic. The most widely used transition metal in biology, iron (Fe), is extremely toxic in its free reduced form (Fe^{+2}) in aerobic environments, as it can catalyze the production of hydroxyl radicals from oxygen using readily available biological reductants. Hydroxyl radicals will indiscriminately react with and destroy essentially any organic molecule they encounter including DNA, causing strand scission. In addition, essentially any element either in concentrations too high and/or in a particular state is

toxic. Further, many "toxic" elements, such as chromium (Cr), arsenic (As), bromine (Br), fluorine (F), and cadmium (Cd), are essential (or likely to be so) for at least some life forms. This brings up the important point that there is great diversity in the elements and nutrients required for growth across the biosphere and what is deadly for some is necessary for others.

ICP-MS is a powerful technique for the detection of elements, and the specific set of elements absorbed by different organisms can be expected to vary significantly (5). Organism or organ/tissue level analysis of all metals assimilated is useful from a variety of perspectives. For instance, it could be used to screen organisms or communities for possible bioremediators, identify the site of metal toxicity or metal accumulation/storage within a multicellular organism, or as a first-pass screen for the isolation of novel metalloproteins or sites of element toxicity.

From the perspective of the protocols presented here, toxicity and use of a given element are both very often the result of a tight association between a macromolecule (generally a protein) and the element of interest. In either case ICP-MS can be used to produce metal elution profiles from chromatographic columns and one can purify metal peaks over multiple column chromatography steps as one might by tracking a peak of enzymatic activity. Thus, the methods described here outline a way to purify metalloproteins based on the metal(s) they contain. This allows for identification of metalloproteins that have hitherto gone unnoticed.

The first protocol presented describes the analysis of complex biological samples by ICP-MS. It is meant to be broad in scope, both in terms of the types of samples that can be analyzed as well as the number of elements that can be analyzed from one sample of limited quantity (less than 1 ml). Support protocol 1 describes the acid washing procedure for ICP-MS sample containers. The second and third protocols describe specific applications of the steps outlined in Basic Protocol 1. Basic Protocol 2 relates to low abundance elemental analysis of microbial

biomass, although it could be adapted to other sources of biomass. Basic Protocol 3 focuses on identification of a novel metalloprotein. Support Protocol 2 explains how to acid wash glassware for use in Basic Protocols 2 and 3 to reduce metal contamination and Support Protocol 3explains how to prepare anaerobic solutions for Basic Protocols 2 and 3 if they are to be carried out anaerobically to protect oxygen sensitive metalloproteins from losing their metal.

BASIC PROTOCOL 1: ICP-MS ANALYSIS OF SOLUBLE PROTEIN SAMPLES

Inductively coupled plasma mass spectrometry (ICP-MS) enables the measurement of a wide range of metals and some non-metallic elements in biological samples to a high degree of accuracy over a wide range of concentrations. ICP-MS has a number of advantages over other techniques for trace metal analysis such as atomic absorption or emission spectroscopies, including high sensitivity coupled with the ability to measure multiple elements over a wide concentration range in a single run. The ability to detect 65 elements below 1 ppt (10) yields useful concentration data with minimal a priori knowledge of a sample's composition, dramatically increasing the throughput of exploratory analyses. ICP-MS achieves its sensitivity and robustness by coupling a high temperature plasma, in which molecular bonds are broken and atoms ionized, to mature mass spectrometric separation technology such as quadrupole/octopole or time of flight allowing sensitive detection of many isotopes. A number of instruments are available from manufacturers such as Agilent and Perkin-Elmer with options tailored to specific fields with unique requirements, particularly in sample introduction techniques. The introduction of biological fluids such as those generated by basic protocols 1 and 2 can be achieved with standard spray chamber/nebulizer configurations. Acid digestion is generally sufficient pre-processing for the types of samples discussed in this protocol, but other methods

such as microwave digestion are routinely applied to less labile samples. Given the sensitivity of elemental detection via ICP-MS, sample requirements are very modest (typically <100µL) which is critical in the analysis of fractions produced by muti-level chromatography as described in Basic Protocol 3, which represent a very small, highly purified fraction of the initial biological material.

To accurately measure trace amounts of metals, it is vital to limit the introduction of metal contaminants to the greatest extent possible. This includes use of high purity reagents, glass distilled water, and acid-washed sample containers. The specifics of instrument operation including tuning, calibration, maintenance, and sample introduction vary greatly among instruments and the operator must tailor analysis procedures according to the manufacturer's recommendations. Appropriate internal and external calibration standards must be chosen for the elements of interest. The techniques and results described in these protocols were performed using an Agilent 7500cs octopole ICP-MS with ShieldTorch and a collision/reaction cell. Herein we describe the analysis of the cytoplasm of a microorganism. This does not present many difficulties in sample preparation and introduction, especially compared to environmental samples that frequently have high suspended solid content. Analysis of a few samples can be accomplished by direct sample introduction, but Basic Protocols 2 and 3 generate large numbers of samples and this necessitates the use of an autosampler. To ensure data reliability, replicate analysis of each sample should be performed.

Materials

Protein sample as generated by Basic Protocols 2 and 3

Agilent 7500ce ICP-MS with Micro-Mist nebulizer

Cetac ASX-520 autosampler

Acid-washed plastic sample containers (15 ml falcon)

Glass distilled deionized (gdd) water

Trace-metal free nitric acid (2% and 5% v/v solutions)

Tuning Solution for ICP-MS 7500cs (Agilent PN: 5185-5959)

External calibration standards

IV-ICPMS-71A (Inorganic Ventures)

CCS-5 (Inorganic Ventures)

CMS-2 (Inorganic Ventures)

Internal Standard

IV-ICPMS-71D (Inorganic Ventures)

Heated shaking incubator

Beckman Allegra 6R Centrifuge

Vortexer

Initial preparation for ICP-MS analysis

1. Start instrument and ensure proper tuning according to manufacturer's recommendations.

2. To prepare the internal standard solution for in-line addition, add 1 ml IV-ICPMS-71D to 9 ml gdd water, mix well and insert the inline addition tube into internal standard tube and place tubing into peristaltic pump. Connect the tubing to the sample introduction block.

3. Create a run method by choosing the isotope masses to be analyzed with the appropriate integration time and acquisition mode. Agilent ChemStation or MassHunter software provides the Offline Acquisition Editor tool to create and modify run methods, including a method creation wizard. The choice of isotopes to be analyzed is governed by the experimental question at hand. For initial analysis of a complex mixture, a large set of elements of interest are typically measured while for terminal fractions containing a purified protein, only a single element of interest may be measured. The default recommended isotopes for elements to be analyzed can be added to a method by clicking on a periodic table interface. To obtain accurate concentration values, data should be acquired using the quantitative analysis mode.

Preparation of standards and creation of external calibration line

4. Prepare a standard curve of the elements of interest by first preparing 1 ppm standard which consists of: 3.95 ml 2% nitric acid, 500 µL IV-ICPMS-71A, 500 µL CMS-2, and 50 µL CCS-5. Mix well.

5. Next prepare the following standard dilutions:

    a. Add 250 µL of 1 ppm standard to 4.75 ml of gdd water (for 50 ppb).
    b. Add 50 µL of 1 ppm standard to 4.95 ml of gdd water (for 10 ppb).
    c. Add 25 µL of 1 ppm standard to 4.975 ml of gdd water (for 5 ppb).
    d. Add 100 µL of 50 ppb standard to 4.90 ml of gdd water (for 1 ppb).
    e. Add 75 µL of 50 ppb standard to 4.925 ml of gdd water (for 0.75 ppb).

f. Add 50 µL of 50 ppb standard to 4.95 ml of gdd water (for 0.5 ppb).

g. Add 25 µL of 50 ppb standard to 4.975 ml of gdd water (for 0.25 ppb).

h. Add 10 µL of 50 ppb standard to 4.99 ml of gdd water (for 0.1 ppb).

i. gdd water (for 0 ppb).

6. Acquire the data and analyze the standard curve on each standard dilution from Step 5 using the method created in Step 3.

7. Create a calibration line for each element measured, using the appropriate internal standard. Inspect each calibration line for acceptable linear fit and %RSD and save the calibration file.

Preparation of soluble protein samples for ICP-MS and data acquisition

8. Prepare the protein samples by thawing (if stored at -80°C) and bringing to room temperature. Mix the samples by spinning briefly in a microcentrifuge and gently vortexing to ensure homogeneity.

9. Add 100 µL of the mixed sample to 1.4 ml of 2% v/v nitric acid in a 15 ml acid-washed falcon tube followed by vortexing and incubation at 37°C for 1.5 hours with shaking @ 250 rpm. Spin each acid-treated sample for 10 minutes @ 3200 rpm (2400 x g) to settle liquid on sides of tube and pellet any precipitate. De-cap tubes and place in autosampler.

10. Create a run sequence by entering the coordinates of each tube in the order of analysis and the method to be used (Step 3). Save the run sequence, and initiate the analysis.

    If the expected concentration of a measured element in a sample is very high, the addition of one or more wash steps with gdH_20 or 2% nitric acid may be desired to eliminate possible carryover effects in subsequent samples

    .

11. After the data acquisition is complete, apply the data to the calibration file created in Step 7 and export the data to a tabular format.

12. Finally, process the data by converting the ppb data to µM (ppb ÷ atomic mass), computing averages and standard deviations, and plot the data (as shown in Basic Protocol 3 and Anticipated Results sections).

BASIC PROTOCOL 2: PREPARATION OF WASHED CYTOPLASMIC EXTRACT FOR THE IDENTIFICATION OF SOLUBLE METAL-MACROMOLECULAR COMPLEXES (≥ 3 kDa)

Sample preparation is the first critical step in any metalloproteomics experiment. It dictates the quality of both the sample and subsequent metal analysis. The following analytical-scale protocol uses a minimum number of steps with the aim of maintaining the integrity of biological metal-macromolecule interactions. Cells from an organism of interest are broken and fractionated in as metal-free an environment as practical and subjected to minimal handling to

obtain a reproducible and accurate estimate of strong and weak metal-macromolecule interactions in the cytoplasmic (or soluble) fraction. To achieve this, the soluble fraction undergoes a series of buffer washes through a centrifugal filtration device (3-5 kDa cut-off). The fractions generated at each step are analyzed by ICP-MS to determine the metal content and thereby the susceptibility of metals to removal from the cytoplasm by simple filtration. Those that are retained by the filter (≥ 3 kDa) cannot be free metal ions and must be bound to macromolecular complexes, which are generally assumed to be proteins. The protocol described here is summarized in Figure 1. It was used to identify metal-macromolecular interactions in the hyperthermophilic archaeon, *Pyrococcus furiosus*, which grows optimally at 100°C (Cvetkovic et al 2010). However, this protocol can be easily adapted by using buffers and cell lysis conditions optimized for any organism of interest.

Materials

Vacuum manifold connected to vacuum pump and argon tank

3 g wet weight of freshly prepared cell pellet or flash frozen *P. furiosus* cells (stored at -80°C)

100 ml anaerobic 50 mM TrisHCl (8.0)

DNase I (5 mg)

10 ml serum bottle with rubber stopper

Anaerobic glove box

Ultracentrifuge (Beckman Optima L-90K)

Fixed angle rotor for ultracentrifuge (Beckman 70.1 Ti: 12 x 10.4 ml capacity)

Polycarbonate centrifuge tubes with cap assemblies (10.4 ml capacity; Beckman 355603)

Tabletop Eppendorf centrifuge 5430 (in chamber)

Eppendorf fixed angle rotor F-35-6-30 with adaptors (holds 6 x 15 ml centrifugal filter devices)

Millipore's Amicon Ultra-4 centrifugal filter devices (3K nominal molecular weight limit)

50 ml glass beaker(s) with magnetic stir bar(s)

Magnetic stirrer (in chamber)

Pipets, pipet tip boxes and corresponding pipettors

25 or 30 ml glass serum bottles with rubber stoppers

5 ml glass serum bottles with rubber stoppers

1.5 ml sterile polypropylene (pp) screw cap tubes (Phenix Research Products, Catalog # SCS-015FS)

Bradford reagent with bovine serum albumin protein standard (Biorad, Catalog #500-0002)

Additional materials for doing Bradford protein assay (see Bio-Rad protein assay manual)

NOTE: All steps are carried out at room temperature unless otherwise stated, although samples from microbes that grow at much lower temperature may require processing at 4°C to preserve the integrity of their proteins.

NOTE: All clean glassware and stir bars used for the experiment should be acid-washed as described in Support Protocol 2. Cleaned centrifuge tubes/lid assemblies should be soaked and rinsed multiple times in glass-distilled deionized (gdd) water or equivalent. Chemicals should be of the highest purity possible. All solutions should be made in gdd water and stored in acid-washed glass containers.

NOTE: All manipulations should be done under strictly anaerobic conditions i.e. using a vacuum manifold and/or in an anaerobic chamber. The protocol described here is for 3 g wet weight of cells – but can be scaled up if desired. Transfer all glassware, stir bars, rubber stoppers, tip boxes, pipettes, centrifuge tubes with caps (open) and centrifugal filtration devices with caps (open) into the anaerobic chamber the evening before starting the experiment. This will allow any diffused oxygen in the plastic/glass/pipet tips/rubber stoppers to be scrubbed out so that everything is oxygen-free by the time the sample is introduced into the chamber.

Preparation of whole cell extract

1. Prepare 100 ml of anaerobic cell lysis/wash buffer (50 mM Tris HCl, pH 8.0) as described in Support Protocol 3. Transfer the buffer into the anaerobic chamber, along with powdered DNAse I (5 mg) placed in a 10 ml serum bottle. Add 5 ml of buffer to the DNAse I to yield a 1 mg/ml stock solution.

   The addition of reductant (sodium dithionite) to the buffer is optional. When reductant is used, it is added to a final concentration of 2 mM after all the oxygen is removed from the buffer.

2. Transfer the fresh cell pellet or frozen cells (3 g) into the anaerobic chamber. If using a fresh cell pellet, resuspend in three volumes (9 ml) of buffer and transfer into a 50 ml beaker with stirring. If starting with frozen cells, drop 3 g of frozen cell pieces into 9 ml of buffer in a 50 ml beaker while stirring. Continue stirring until there are no visible clumps of cells and the extract is homogeneous (~ 30 mins).

P. furiosus undergoes osmotic lysis when placed in the hypotonic lysis buffer. The release of cellular DNA makes the extract very viscous. This viscosity can be used as a criterion to gauge cell lysis.

3. Add 4.5 µl of the DNAse I stock solution (1 mg/ml) to the lysed cells and continue stirring until the whole-cell extract is free-flowing (~ 30 - 60 mins).

The change in viscosity indicates that the DNAse I has digested the DNA in the extract.

Preparation of cytoplasmic extract (S100*)*

4. Transfer the cell extract into a 10.4 ml ultracentrifuge tube and carefully seal the tube using the O-ring and screw cap.

5. Bring the sealed sample out of the anaerobic chamber and ultracentrifuge at 15°C for 1 hr at 100,000 x *g* (38,200 rpm in a Type 70.1 Ti rotor).

P. furiosus is a hyperthermophile and has thermostable biomolecules. The temperature setting used during centrifugation of whole-cell extracts should be optimized for the organism(s) of interest.

6. After centrifugation, gently transfer the sealed sample-containing centrifuge tube back into anaerobic chamber, carefully unscrew the cap and pipet the 100,000 x *g* supernatant

(S100) into a 25 ml acid-washed serum bottle leaving a small amount of liquid on the pellet (to avoid mixing of the pellet with supernatant). Gently swirl serum bottle containing the S100 until well mixed and cap with a rubber stopper.

The S100 represents the unwashed cytoplasmic fraction.

Washing the cytoplasmic extract (S100) using an Amicon Ultra-4 centrifugal filter device

7. Wash the Amicon Ultra-4 filter device with 2 ml of buffer at 4,300 x *g* (6,600 rpm) until almost empty. Do not allow the membrane to dry out. Remove the buffer from the centrifuge tube holding the filter device and any buffer remaining in the filter device and repeat the wash. Remove the buffer from the centrifuge tube holding the filter device and any remaining buffer in the filter device. Rinse out the centrifuge tube with an additional 2 ml of buffer and remove completely.

   The membrane contains traces of glycerine and pre-rinsing with wash buffer removes the glycerine and any additional unbound contaminants.

8. Add 2 ml of the S100 into the filter device and spin at 4,300 x *g* for 10-15 min or until the volume of the S100 is 0.5 ml or a little less. Pipet the flow-through (FT) from the centrifuge tube into a pre-weighed 5 ml acid-washed serum bottle (with stopper) and mix well. If the volume of the concentrated S100 is less than 0.5 ml, bring the volume of the concentrated S100 to 0.5 ml by adding back some of the FT fraction. Stopper the bottle.

The FT represents the same background of small molecules as that present in the S100. By maintaining the minimum volume at 0.5 ml, a four-fold buffer exchange will be achieved with each wash step.

9. Add 1.5 ml buffer to the concentrated S100 to bring the volume to 2 ml. Mix the sample by gently pipeting up and down until completely homogeneous and centrifuge as described in Step 8. Transfer the FT, which represents the first buffer wash (W-1), into a pre-weighed 5 ml acid-washed serum bottle (with stopper) and mix well. Once again, bring the volume of the washed S100 to 0.5 ml with W-1 and mix well.

   Mixing the flow-through, washes and S100 at each step is critical to ensuring homogeneity and accurate ICP-MS measurements. In addition, the centrifugation results in a gradient of protein concentration forming in the filter device that needs to be mixed at each step or it can lead to loss of protein due to precipitation.

10. Repeat Step 9 two more times to generate two additional washes (W-2 and W-3) and 0.5 ml of the final S100 fraction. Add 1.5 ml of buffer to the S100 and mix well to generate the final 256-fold buffer-exchanged, washed S100 (S100w).

    Following this procedure leads to six fractions, each in a stoppered acid-washed glass serum bottle: S100, FT, W-1, W-2, W-3 and S100w. The FT and washes will be 1.5 ml each, while the S100w will be 2 ml. Weigh the stoppered, labeled serum bottles before

and after the addition of each fraction. The difference in weight will yield the exact amount of each fraction (assuming 1g = 1 ml).

Sampling for ICPMS and protein estimation

11. Transfer 0.5 ml of buffer and each sample (S100, FT, W-1, W-2, W-3 and S100w) into individual 1.5 ml pp screw cap vials. The buffer will serve as the ICP-MS blank for the experiment. Store at 4°C for immediate ICP-MS analysis. If the metal analysis is not being done immediately, the samples and fractions can be stored at -80°C for extended periods of time. Use 100 µl of the buffer and an equivalent volume of each fraction (in triplicate) for ICP-MS analysis (53 elements). The samples should be loaded into the ICP-MS in the following order (typically from lower to increasing concentrations of elements): Buffer (blank), W-3, W-2, W-1, FT, S100w and S100.

12. Determine the protein concentration of the S100 and S100w sampled fractions using the Bradford micro-assay as described by the manufacturer (Biorad). The resuspension/lysis buffer will serve as the blank for the assay.

    The protein concentrations in the S100 and S100w should be within the range of the standard deviation of the two samples. If there is substantially less protein in the S100w, that could indicate that the membrane was compromised during the procedure or that unrecovered protein precipitate was left on the filter unit. To distinguish between these two possibilities, test for protein in the FT and wash fractions. If there is protein in these fractions, the experiment will have to be repeated using a new ultrafiltration device

starting at Step 7. If there is no protein in the FT and wash fractions, that could indicate unrecovered protein precipitate left in the filter unit.

The final ICP-MS data from Basic Protocol 1 are expressed as metal concentrations (in units of µM). For a single sample set, the data can be presented as the total amount of each element in each sample (S100, FT, W-1, W-2, W-3 and S100W) in µmoles. For this, one needs to determine the volume of each sample by weighing the serum bottle + stopper before and after adding the sample to calculate the actual volume (assuming 1g = 1 ml) and the total µmoles for each element at each step and present it in the form of a bar graph (see Results section). To compare across sample types, the common thread will be the amount of metal detected per milligram protein in each sample. When using protein as the common denominator, it is important to express the metal data for all the wash steps as µmoles (or nmoles) per milligram of protein - using the protein concentration determined for the starting S100.

<u>Example 1: To determine the total amount of metal in a sample</u>

Weight of empty serum bottle and stopper = 2.5 g

Weight of empty serum bottle + stopper+sample = 4.05 g

Metal concentration of sample = 0.2 µM

Total volume of sample = 4.05 g – 2.5 g = 1.55 ml (assuming 1g = 1 ml)

Total metal in sample = (1.55 ml) x (0.2 µM) = 1.31 µmoles

<u>Example 2: To determine the amount of metal per mg protein in the sample</u>

Metal concentration of sample = 0.2 µM

Protein concentration of sample = 12 mg/ml

Metal content per mg protein = (0.2 µM)/(12 mg/ml) = 0.0166 µmoles metal/mg protein or 16.6 nmoles metal/mg protein

BASIC PROTOCOL 3: ANAEROBIC PURIFICATION OF METALLOPROTEINS USING ICP-MS AND CONVENTIONAL PROTEIN CHROMATOGRAPHY

The techniques described here enable the anaerobic purification and identification of a metalloprotein. The steps given were used to identify a new molybdenum-containing protein from *Pyrococcus furiosus* (Pf), a hyperthermophilic archaeon thought to have no use for molybdenum before molybdenum was found in the cytoplasm via Basic Protocol 2 (5, 18). However, the protocol can be adapted to any protein sample that can be obtained in sufficient amounts including membrane proteins, proteins from a tissue sample or even proteins from a sample of purified organelles. Of course, significant adaptation will be necessary for different cell/sample types and each purification procedure will be unique to the particular protein of interest and will be dependent upon the properties of the target protein as well as its source. Selection of the original protein pool depends on the interests of the researcher and the availability of sufficient sample, and procedures for obtaining biomass will not be covered here. Large quantities of sample (on the order of grams to tens of grams of total protein) are preferred if not required due to significant losses intrinsic to multiple levels of column chromatography as well as to allow for the detection of proteins with rather low abundance and/or sub-stoichiometric metal content. In addition, protein purification by this method is never an exact science and will likely require more than one attempt.

Finally, growth conditions of the target organism should be considered. ICP-MS analysis of the growth medium (for prokaryotes) is useful to determine exactly how much of each element is available for the growth of the organism. Inevitably, many more elements are present in the medium as contaminants of the various chemical (e.g., inorganic salts) and biological components (e.g., yeast extract, peptone) than the elements intentionally added. Increasing the concentration or availability of a given element can have the effect of increasing the levels detected either through a shift in equilibrium or upregulation of the expression of key genes in response to the availability of an element.

Materials

Vacuum manifold (connected to mechanical vacuum pump and argon tank)

Glassware & accessories

15 L carboy that can hold a vacuum x2

4 L sidearm flasks x3

1 L sidearm flasks x2

250 ml sidearm flasks x2

5-125 ml serum bottles/vials

rubber stoppers to fit all flasks, bottles, and vials

plastic/rubber hosing

clamps for sealing off rubber hosing

hose barb connectors

Stir bars

Stir plate

Syringes & needles

Buffers and salts (Tris, K_3PO_4, NaCl, $(NH_4)_2SO_4$, HCl, NaOH)

Glass distilled deionized water

1.1 L DEAE Sepharose Fast Flow (FF) column (GE product #17-0709-05, column - GE XK 50/60)

Superdex 75 26/60 (320 ml) size exclusion column

8 ml hydroxyapatite column (Ceramic hydroxyapatite - Bio-Rad CHT Type I, 20 µm particle ceramic Hydroxyapatite, product #158-2000, column - GE XK 16/20)

1 ml Phenyl sepharose column (e.g. GE HiTrap Phenyl HP)

1 ml MonoQ 5/50 GL column

Protein chromatography system (e.g. AKTA BASIC)

Anaerobic glove box

300 g frozen fermenter-grown Pf cells

DNase I

Ultracentrifuge with rotors and tubes

Electrophoresis apparatus and SDS-PAGE gels

Bradford protein assay kit (e.g. Bio-Rad #500-0201)

Advantec 450 ml stirred pressure filtration cell (Model # UHP-76)

Millipore 76 mm Ultrafiltration membrane (3kDa nominal molecular weight limit) (Cat. # PLBC07610)

Obtain soluble cell-free extract

1. Prepare 900 ml anaerobic 50 mM Tris pH 8.0 (Buffer A) according to Support Protocol 3 using acid-washed glassware (Support Protocol 2).

2. Add 300 g frozen Pf cells and 25 mg DNase I to buffer prepared in Step 1.

3. Stir and continue to exchange atmosphere with inert gas on a vacuum manifold to remove residual oxygen until sample has thawed (about 30 min.).

4. Continue to stir until sample is homogenous and has low viscosity (visually similar to water).

   P. furiosus cells are broken by freeze-thaw and osmotic shock.

5. Take the sample and ultracentrifuge tubes into a glove box.

6. Add the crude cell extract to the centrifuge tubes, balance, and load into an ultracentrifuge

7. Spin for 1 hour at 100,000 x *g* (rcf) at 15°C.

8. Take the spun tubes back into the glove box, pool the supernatants in a flask, and seal. Repeat Steps 6-8 until all cell extract has been centrifuged.

Column 1: anion exchange (DEAE FF) chromatography

9. Prepare 25 L anaerobic buffer A and 5 L anaerobic 2 M NaCl, 50 mM Tris-HCl, pH 8.0 (buffer B) according to Support Protocol 3 using acid-washed glassware (Support Protocol 2).

   Since Pf is a hyperthermophilic organism, its proteins are quite stable at room temperature. For separation of mesophilic extracts, it may be best to cool buffers to 4°C to reduce protein denaturation and precipitation.

10. Wash the 1.1 L DEAE column (C1) with 3 column volumes (CV) water and then 3 CV buffer A at 20 ml/min (continue to use this flow rate for C1).

    For guidance on assembling a column from fresh column support material and an empty column, refer to (3). Buffers and samples should be loaded onto the chromatography system through needles to exclude oxygen. Argon flow through the flask should be maintained and the flask vented through a bubbler to maintain anaerobicity and atmospheric pressure. See Figure 2 for example.

11. Retain 3 ml of the flow-through buffer A to use as an ICP-MS buffer blank.

    Other useful ICP-MS control samples are buffer A and buffer B taken directly from the flask before they are applied to the column.

12. Load the soluble cell extract onto C1 as a 30:70 mix of extract:buffer A, collecting the flow-through either through a needle into an anaerobic flask vented through a bubbler or with the outlet of the chromatography system running into a glove box.

13. Wash C1 with 3 CV buffer A and collect the wash anaerobically.

14. Run a gradient from 0-25% buffer B (100-75% buffer A) in 15 L and collect 125 ml fractions anaerobically (120 fractions).

15. Run a gradient from 25-100% buffer B (75-0% buffer A) in 3 L and collect three 1 L fractions anaerobically.

16. Retain 3 ml of the flow-through buffer B to use as an ICP-MS buffer blank.

    This should be done after conductivity and UV-visible traces are completely flat indicating that the column is clean and pure buffer B is flowing through.

17. Using a syringe and needle, take 1 ml from each eluate sample without exposing it to oxygen. These 1 ml samples need not be kept anaerobic. For each one:
    a. Measure molybdenum concentration in duplicate via ICP-MS according to Basic Protocol 1.
    b. Measure protein concentration via Bradford assay.

c. Run SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis). For additional guidance, consult (8).

Example data for molybdenum and protein concentrations from a first-level (C1) DEAE column are shown in Figure 3. It may be desirable to measure multiple or even many metals especially on earlier columns in the purification, but only molybdenum is followed here.

18. Using all data collected, select fractions and pool them. In Figure 3, the PF1972 peak is centered near fraction 81.

The goal is to select one peak of a given metal with the possible exclusion of fractions which may introduce additional contaminants in significant quantities (as revealed by SDS-PAGE). Metal to protein molar ratio is a useful metric when deciding which fractions to pool for the next column and can be used track increasing purity over multiple columns. See Figure 3 for example data in which an appropriate fraction pool is boxed.

Column 2: size exclusion (superdex 75) chromatography (all steps done anaerobically unless otherwise noted)

19. Concentrate fraction pool from Step 18 to 5 ml using the stirred pressure filtration cell pressurized with argon and 3 kDa NMWL membrane.

20. Prepare 1.2 L anaerobic 50 mM Tris, 300 mM NaCl, pH 8.0 (buffer C) according to Support Protocol 3 using acid-washed glassware (Support Protocol 2).

21. Equilibrate the Superdex 75 26/60 (320 ml) column (C2) with 1 CV water and then 2 CV buffer C at 2 ml/min (continue to use this flow rate for C2).

22. Continue flow and retain 3 ml of the flow-through buffer C to use as an ICP-MS buffer blank.

23. Load the fraction pool concentrate from Step 19 onto a 5 ml loop through a syringe.

24. Inject the sample onto the column by directing buffer through the sample loop.

25. Collect two 50 ml fractions.

    These two fractions contain the void volume of proteins too large to be separated by the column (>70 kDa).

26. Collect forty 5 ml fractions.

    These fractions contain high molecular weight species (3-70 kDa), fractionated according to decreasing size.

27. Collect two 10 ml fractions.

    These fractions contain lower molecular weight species(<3 kDa), which are not separated by the column.

28. Using a syringe and needle, take 0.4 ml from each eluate sample without exposing it to oxygen. These samples need not be kept anaerobic. For each one:
    a. Measure molybdenum concentration in duplicate via ICP-MS according to Basic Protocol 1.
    b. Measure protein concentration via Bradford assay.
    c. Run SDS-PAGE.

    Example data for molybdenum and protein concentrations from a second-level (C2) size exclusion column are shown in Figure 4.

29. Using all data collected, select fractions to pool. In Figure 4, the PF1972 peak is centered near fraction 20.

    See Figure 4 for example data in which an appropriate fraction pool is boxed.

Column 3: hydroxyapatite chromatography (all steps done anaerobically unless otherwise noted)

30. Prepare 600 ml anaerobic 5 mM potassium phosphate, pH 8.0 (buffer D) and 100 ml anaerobic 500 mM potassium phosphate, pH 8.0 (buffer E) according to Support Protocol 3 using acid-washed glassware (Support Protocol 2).

31. Wash the 8 ml hydroxyapatite column (C3) with 3 CV water and then 3 CV buffer D at 4 ml/min (continue to use this flow rate for C3).

32. Retain 3 ml of the flow-through buffer D to use as an ICP-MS buffer blank.

33. Load the fraction pool from Step 29 onto C3 as a 5:95 mix of sample:buffer D collecting the flow-through.

34. Wash C3 with 100 ml buffer D and collect the wash.

35. Run a gradient from 0-50% buffer E (100-50% buffer D) in 80 ml and collect forty 2 ml fractions.

36. Run a gradient from 50-100% buffer E (50-0% buffer D) in 12 ml and collect four 4 ml fractions.

37. Collect a 36 ml wash at 100% buffer E.

38. Retain 3 ml of the flow-through buffer E to use as an ICP-MS buffer blank.

39. Using a syringe and needle, take 0.35 ml from each eluate sample without exposing it to oxygen. These samples need not be kept anaerobic. For each one:
    a. Measure molybdenum concentration in duplicate via ICP-MS according to Basic Protocol 1.
    b. Measure protein concentration via Bradford assay.
    c. Run SDS-PAGE.

    Example data for molybdenum and protein concentrations and from a third-level (C3) hydroxyapatite column are shown in Figure 5.

40. Using all data collected, select fractions and pool them. In Figure 5, the PF1972 peak is centered near fraction 25.

    See Figure 5 for example data in which an appropriate fraction pool is boxed.

Column 4: hydrophobic interaction (phenyl sepharose) chromatography (all steps done anaerobically unless otherwise noted)

41. Prepare 50 ml anaerobic 50 mM potassium phosphate, 2 M $(NH_4)_2SO_4$, pH 7.0 (buffer F) and 50 ml anaerobic 50 mM potassium phosphate, pH 7.0 (buffer G) according to Support Protocol 3 using acid-washed glassware (Support Protocol 2).

42. Wash the 1 ml phenyl sepharose column (C4) with 3 CV water and then 3 CV buffer F at 1 ml/min (continue to use this flow rate for C4).

43. Retain 3 ml of the flow-through buffer F to use as an ICP-MS buffer blank.

44. Load the fraction pool from Step 40 onto C4 as a 50:50 mix of sample:buffer F collecting the flow-through.

45. Wash C4 with 6 CV 50% buffer G (50% buffer F) and collect the wash.

46. Run a gradient from 50-100% buffer G (50-0% buffer F) in 20 ml and collect twenty 1 ml fractions.

47. Collect ten 1 ml washes at 100% buffer G.

48. Retain 3 ml of the flow-through buffer G to use as an ICP-MS buffer blank.

49. Using a syringe and needle, take 0.35 ml from each fraction without exposing it to oxygen. These samples need not be kept anaerobic. For each one:
    a. Measure molybdenum concentration in duplicate via ICP-MS according to Basic Protocol 1.
    b. Measure protein concentration via Bradford assay.
    c. Run SDS-PAGE.

Example data for molybdenum concentrations from a fourth-level (C4) phenyl sepharose column are shown in Figure 6. Protein concentrations were too low for quantification.

50. Using all data collected, select fractions and pool them. In Figure 6, the PF1972 peak comes out in the low salt (100% buffer G) washes.

    See Figure 6 for example data in which an appropriate fraction pool is boxed. Due to the low protein concentration in these fractions, the protein was not measured directly.

Column 5: anion exchange (MonoQ) chromatography (all steps done anaerobically unless otherwise noted)

51. Prepare 100 ml anaerobic 50 mM Tris, pH 8.0 (buffer H) and 50 ml anaerobic 50 mM Tris, 1 M NaCl, pH 8.0 (buffer I) according to Support Protocol 3 using acid-washed glassware (Support Protocol 2).

52. Wash the 1 ml MonoQ column (C5) with 3 CV water and then 3 CV buffer H at 1 ml/min (continue to use this flow rate for C5).

53. Retain 3 ml of the flow-through buffer H to use as an ICP-MS buffer blank.

54. Load the fraction pool from Step 50 onto C5 as a 20:80 mix of sample:buffer H collecting the flow-through.

55. Wash C5 with 50 ml 20% buffer I (80% buffer H) and collect the wash.

56. Run a gradient from 20-40% buffer I (80-60% buffer H) in 40 ml and collect forty 1 ml fractions.

57. Collect five 1 ml washes at 100% buffer I.

58. Retain 3 ml of the flow-through buffer I to use as an ICP-MS buffer blank.

59. Using a syringe and needle, take 0.2 ml from each eluate sample without exposing it to oxygen. These samples need not be kept anaerobic. For each one:
    a. Measure molybdenum concentration in duplicate via ICP-MS according to Basic Protocol 1.
    b. Measure protein concentration via Bradford assay.
    c. Run SDS-PAGE.

Example data for molybdenum and protein concentrations and SDS-PAGE of protein elution from a fifth-level (C5) phenyl sepharose column are shown in Figure 7. Due to the low protein concentration in these fractions, the protein was visualized using a silver stain.

60. Submit the single gel band for MALDI analysis at a mass spectrometry facility.

SUPPORT PROTOCOL 1: ACID-WASHING: PLASTICWARE FOR ICP-MS SAMPLE HANDLING

Support protocol 1 is for cleaning the plasticware that is used for ICP-MS sample processing whereas Support protocol 2 is used for cleaning the glassware that is used in the biological protocols i.e cell breakage, S100 wash samples and fractions generated during purification using column chromatography. Plastic and glassware presents a major source of metal contamination in processing biological samples. The use of nitric acid solutions in sample processing and as the final matrix for sample introduction can cause metals to be released from tubes. It is important to pre-treat sample vessels with nitric acid and rinse away the leached metal. Trace metal grade acid and glass distilled water which has not contacted any metal is essential in reducing metal contamination.

Materials

Three high density polyethylene (HDPE) tanks (Norton)

2% v/v nitric acid

gdd water

Large Kim-wipes

New 15 ml screw cap conical polypropylene tubes (Corning cat. No. 430052)

Protocol steps

1. Place all plastic tubes which will contain samples for ICP-MS into 2% v/v Nitric acid solution tank.

2. After 24 hours, transfer plastic tubes into clean glass-distilled water tank.

3. After 24 hours, transfer plastic tubes into second clean glass-distilled water tank.

4. Remove plastic tubes and place them inverted onto clean kim-wipes to allow drying. Tubes can be capped for storage when drying is complete.

SUPPORT PROTOCOL 2: ACID-WASHING: GLASSWARE FOR ANAEROBIC CYTOPLASMIC WASHES AND PURIFICATION OF METALLOPROTEINS

A key requirement for maintaining biological metal-macromolecular interactions is an anaerobic environment. Plastics are porous and contain trapped air that is released slowly over time in an anaerobic environment. To minimize the introduction of air, all anaerobic work is done using glassware, which is impervious to air. Acid-washing the 'clean' glassware that will come in contact with samples being processed for ICP-MS analysis will minimize contamination with non-biological metals that may be present in the glassware.

Materials

Polypropylene tank with lid (Nalgene (27L), Fisher Scientific # 14-831-112)

Polypropylene tray (Nalgene (15L), Fisher Scientific #13-359-20 B)

Deionized water

gdd water or equivalent purity water

Reagent grade nitric acid (Sigma-Aldrich #438073)

Long acid-resistant polyethylene or latex gloves

Clean glassware to be acid-washed

Clean teflon-coated stir bars to be acid-washed

Preparation of nitric acid bath

1. Ensure you are wearing the proper personal protective equipment when handling acid. Wear goggles or a face shield, acid-resistant long rubber latex gloves, lab coat, long pants and closed toed shoes.

2. Rinse the 27-L polypropylene tank (with lid) several times with deionized water, drain and place into a heavy duty 15-L Nalgene polypropylene tray.

3. Fill the tank with 19.6 L of deionized water and slowly add 400 ml of reagent grade nitric acid with gentle stirring (you can use a long plastic coated magnetic rod) to a final concentration of 2% v/v.

Cleaning the glassware

4. All starting glassware and stir bars should be physically clean and free of both chemical residue and grease.

5. Rinse the glassware thoroughly by filling completely with deionized water and discarding water. Drain well.

6. Place the rinsed, drained glassware into the 2% nitric acid bath, making sure that each item is completely submerged i.e. such that all surfaces are in contact with the acid and there is no trapped air. Soak for at least 16 hr (overnight is convenient).

7. Remove the glassware after draining out as much of the acid as possible and place into a clean polypropylene tray. Rinse the outside of the glassware and then fill and drain each piece at least 4 to 5 times with deionized water. Measure the pH of the inside of several random pieces of glassware using pH paper to make sure the pH is not acidic.

8. Rinse the glassware twice by filling part way with gdd water and shaking to distribute the water before discarding. Drain well and dry inverted. Store in closed boxes (inverted) or cover all openings with plastic wrap/screw cap closures.

Note: Some items e.g. large flasks, sidearm flasks, carboys, measuring cylinders etc, may be too tall or awkward for soaking in the acid bath. Such items can be processed as described in Steps 4 and 5 and, instead of soaking in acid, filled with acid, sealed with plastic wrap, labeled with the appropriate hazard label and placed in a clean heavy duty polypropylene tray in a general purpose fume hood. Soak overnight and continue with Step 7.

SUPPORT PROTOCOL 3: PREPARING ANAEROBIC SOLUTIONS

Anaerobic techniques, like sterile techniques, require the development of a certain mindset. Every possible source of air contamination must be suspected. This includes every solution the sample comes in contact with, every joint (e.g. between flask and stopper, stopper and needle, rubber hosing connections etc.), and over the long term, every piece of plastic or rubber, as these materials are gas permeable. It takes only a very small amount of air to ruin an oxygen-sensitive sample especially at the low concentrations involved in the experiments described here. Due to the large volume of buffer that proteins are exposed to during chromatography, it is particularly important to have completely oxygen-free chromatographic buffers. Techniques for removing oxygen from such solutions are given here.

Materials

Solution (e.g. buffer)

Stir bar

Stir plate

Vacuum manifold

Rubber stoppers to fit container (sidearm Erlenmeyer flask)

Soft rubber septa for solution removal

Plastic/rubber hosing

Clamps for sealing off rubber hosing

Hose barb connectors

Protocol steps

1. Place a stir bar in the container of the aerobic solution.

2. Seal the container with a stopper.

   A hard rubber stopper with a single hole plugged with a softer rubber septum (such as those used for stoppered vials) is convenient for running buffer anaerobically onto a column.

3. Connect the container to the vacuum manifold.

   For sidearm flasks, the sidearm can be fitted with a piece of hosing. A clamp (often called hosecock or pinchcock) will be needed to seal off the hosing before disconnecting from the manifold. The hosing can be connected to the manifold directly or to dedicated manifold tubing through a coupling connection. Smaller containers can be connected via a needle through the septum, but this slows the process of removing oxygen due to the flow restriction. Generally a syringe can be fitted to the end of tubing connected to the manifold for easy connection to the needle.

4. Turn on stir plate at medium speed.

   Stirring facilitates gas exchange and prevents bumping (violent boiling).

5. Make sure all manifold valves are closed, turn on the vacuum pump, and open the container to vacuum via the manifold valve.
6. Allow container to stir under vacuum for 5 min.

The effectiveness of atmosphere exchange is related to the surface area to volume ratio. Longer times may be required for large volumes (especially 15 L carboys).

7. Fill container with argon.
8. Repeat Steps 6-7 four to seven times.

More cycles are needed for larger volumes especially those with a lower surface area to volume ratio.

REAGENTS AND SOLUTIONS

Cell lysis/wash buffer *(50 mM Tris·Cl, pH 8.0)*

Materials

1M TrisHCl (pH 8.0) made in gdd water stored in acid-washed glass bottle

Glass-distilled, deionized water (gdd water)

250 ml glass beaker with magnetic stir bar

100 ml glass measuring cylinder

250 ml Pyrex storage bottle with 45 mm screw cap

125 ml Millipore Steritop disposable 45 mm bottle top filter (0.22 µM)

250 ml side arm flask with magnetic stir bar and rubber stopper

Dilute 5 ml of 1 M Tris HCl (pH 8.0) into 95 ml of gdd water in a 250 ml acid-washed glass beaker with stirring and re-adjust to pH 8.0. Take a 125 ml, 0.22 µM bottle top filter attached to a 250 ml screw cap bottle and wash the filter (using vacuum) with gdd water. Drain well. Transfer the washed filter unit onto a 250 ml acid-washed bottle and vacuum-filter the buffer. Transfer the filtered buffer into an acid-washed 250 ml sidearm flask with a stir bar and make anaerobic (with stirring) by going through several cycles of vacuum/argon on a vacuum manifold (see Support protocol 3 for details). Prepare fresh on the day of the experiment.

COMMENTARY

Background Information

These protocols apply ICP-MS to analysis of low-abundance, acid-labile elements in biological samples. This highly sensitive technique is under active development as a part of analytical chemistry and has been widely applied to many fields. ICP-MS has been applied to biological samples since the 1980s, most often to measure the amount of a certain metal present in a given biological sample or to quantify the metal bound to a purified metalloprotein. The former is comparable to Basic Protocol 2, which identifies the metals present in the metallome of *Pyrococcus furiosus* and further determines how much of each element is tightly bound in a form >3kDa (presumably protein-bound).

Basic Protocol 2 was developed to provide a simple way to identify biological metal-macromolecule interactions. The cytoplasmic fraction of any organism grown under a variety of conditions can be subjected to a series of buffer exchanges through a low molecular weight cut off membrane (3 kDa) that separates unbound or weakly-bound metals (< 3kDa) from tightly-bound metal-macromolecules (> 3kDa). By analyzing the original extract, the final washed

extract and all the intermediate washes, one can determine the metals that the organism assimilates from the growth media, and, which of those metals are strongly associated with biological macromolecules (>3 kDa). A comparison of the metal profiles generated for different organisms and/or a single organism grown in different media types can determine species-specific metal or growth-specific metal assimilation patterns. The relative concentration of the various metals (expressed per mg of protein) indicates the *in vivo* abundance of the tightly-bound metals. This method has been used to identify strong and weak metal interactions in the cytoplasmic fraction of *Pyrococcus furiosus* (5).

Basic Protocol 3 combines conventional chromatographic protein purification with ICP-MS as a detection method to purify a protein that contains a specific metal of interest. Protein chromatography is a well-developed field with abundant literature, resources, and commercial products (6). Different column matrices are used to separate proteins based on different physical properties of the molecules in the mixture. A detailed description can be obtained from the column manufacturer, but a brief overview of column types is given here. Ion exchange columns interact with charges/charged regions on the protein surface. Anion exchange materials bind to negatively charged residues on the surface, and thus proteins with more exposed negative residues will bind more tightly and a higher pH will tend to increase binding. The eluent is an anion, such as Cl^- from NaCl, see (19). Phenyl sepharose interacts with hydrophobic residues/regions of the protein surface. Due to the hydrophobicity of the protein interior, phenyl sepharose can cause some proteins to become structurally unstable when they bind. Binding is facilitated by a high concentration of so-called chaotropic agents that promote hydrophobic interactions, and proteins are eluted by decreasing the concentration of these chaotropic agents in a reverse gradient (high to low), see (12). Hydroxyapatite is a calcium phosphate hydroxide

mineral and phosphate is used as the eluent. The affinity of protein molecules for the column matrix is a combination of interactions between negatively charged residues and calcium ions, positively charged residues and phosphate, and hydrogen bonding interactions between phosphate or hydroxide moieties on the column and protein, see (3). Size exclusion (also known as gel filtration or gel permeation) chromatography separates species based on hydrodynamic radius, which can be correlated to molecular weight for globular proteins. Size exclusion supports are highly porous with a variety of pore sizes. The largest molecules cannot enter these pores and quickly pass through the column by flowing around the chromatographic beads whereas smaller molecules enter the pores and their progress is slowed, thus large molecules elute first. A single type of buffer is used (no gradients) and proteins do not actually interact directly with the column material other than entering and leaving the pores. In fact, a high concentration of salt is often used (300 mM NaCl) to help prevent interactions between the column supports and proteins, which would obscure the size-determination, see (11).

CRITICAL PARAMETERS

ICP-MS

Appropriate blanks for background subtraction

While it is vital to limit metal contamination to the greatest extent possible, the complete elimination of trace metal contamination is impossible. It is therefore necessary to analyze all solutions that are used in preparation of samples. This is important both for identification of contaminants that may be problematic, and for subtracting background values to determine the true elemental composition of a sample. If unacceptably high levels of contaminants are

observed in background solutions, it will be necessary to identify and mitigate the source of contamination and then reproduce and reanalyze the samples.

Replicates

Performing technical replicates is recommended for all analyses to ensure acceptable data reproducibility and to identify situations where instrument performance is degraded requiring routine maintenance.

Sample/matrix components

While the types of samples discussed in this protocol do not generally contain as many problematic components as some, such as environmental samples which can contain large quantities of dissolved solids, consideration must be given to ICP-MS sample matrix composition. A variety of strategies for reducing problematic matrix effects have been reported (1). A primary consideration in samples generated by protocol 3 is salt content. Ion exchange chromatography yields fractions of varying salt concentrations, and appropriate dilution must be made to reduce deposition onto sampling and skimmer cones while maintaining sensitivity.

Metal contamination

Care must be taken throughout all procedures to reduce the introduction of contaminant metals as much as possible. Certain transition metals, such as iron, zinc, and copper, are (relatively) abundant in water sources, on surfaces, on skin, as contaminants in salts, reductants, and other chemicals, etc. Thus, it is best to take precautions against contaminating metals not only to eliminate background ICP-MS signals, but also because these can bind non-specifically

to many proteins and lead to false indications that a certain protein binds a given metal *in vivo*. Such precautions include using acid-washed glassware and glass-distilled deionized water, buying chemicals with low contaminant specifications, and being aware that certain metals and plastics will leach metal ions into solutions that they come in contact with.

Anaerobicity

Sample handling is an important consideration. In the majority of cases, it is necessary to conduct all procedures under anaerobic conditions to prevent the loss of metals due to reaction with air (oxygen). This is particularly true with redox-active cofactors that may bind only in a reduced state and dissociate when oxidized. For some more narrowly focused applications, such as the identification of a protein binding a particular oxygen-tolerant metal cofactor, aerobic purification may suffice. However, the authors recommend using anaerobic purification initially, especially if the characteristics of the specific target metal/cofactor are not already known, to ensure that the results convey an accurate picture of the distribution of metals actually present in the sample. In addition to running columns with anaerobic buffer and collecting fractions under anaerobic conditions, it may be wise to use a reductant in all buffers used. This decision must be made carefully because although a reductant can protect metal centers from damage from oxygen or reactive oxygen species, the reduced state of a metal cofactor is often more reactive towards such oxidizing species and can be less stable or bind with lower affinity. Once a protein has been purified, or even if the purification achieves a metallopure peak (only one metal present), oxygen exposure experiments with monitoring of the UV-visible spectrum can show whether or not aerobic purification can be trusted. Any aerobic steps prior to this lead to the risk of losing

metal cofactors, and even if the metal of interest remains, the containing protein may also contain other more oxygen sensitive cofactors.

Chromatographic Purification

The types and sequence of columns that will be used in the purification clearly have an impact on the success of the purification, and although each purification scheme is unique and trial and error plays an important role, there are a few useful guidelines. The authors have most often used an anionic exchange column as the first level column for several reasons, including good resolution, the ability to apply high volume samples (which are concentrated on the column), and the fact that a high proportion of proteins typically bind to the column. Overall, ion exchange, phenyl sepharose (hydrophobic interaction), hydroxyapatite (mixed mode), and size exclusion (gel filtration) columns were most commonly used. The order listed is a good starting point, but will likely require refinement.

TROUBLESHOOTING

ICP-MS & minimizing metal contamination

Troubleshooting specifics vary greatly between instruments, but the most common difficulties presented are signal degradation due to deposition on instrument components, such as extraction cones, skimmer cones and lenses, air introduction due to degraded connections between tubing, tubing blockages, and nonconstant sample introduction rate caused by degraded tubing or peristaltic pump components. Some sample components, such as high salt, can present problems both in the form of signal suppression from non-spectroscopic matrix interferences and deposition onto cones and lenses of the instrument. The most pernicious source of error is

contamination with metals exogenous to the sample. The use of acid-washed glass and plastic vessels, gdd water and high purity contaminant free reagents can mitigate contamination, but when investigating metals at trace levels, a clean dust-free environment and careful pipetting technique is crucial.

Loss of protein during Basic Protocol 2

There are two major causes leading to reduced protein in the final S100w sample: a compromised membrane and/or protein precipitation. Thus it is important not to damage the membrane in the process of adding buffer to the sample, mixing the sample before each spin and removal of the sample from the filter unit. Likewise, mixing the sample and buffer until the sample is homogeneous will greatly minimize the risk of protein precipitation due to protein concentration gradients generated as a result of each centrifugation step.

Chromatographic Purification

The authors have used this protocol to identify new metalloproteins from soluble cytoplasmic extract from cells of the hyperthermophilic archaeon, *Pyrococcus furiosus* (Pf). This is a relatively simple organism in that it contains just over 2000 genes and yet five to seven chromatography steps were required to determine which protein was binding the metal of interest (5). Given the complexity of such mixtures, it can be very difficult to isolate one component. One column of each column type should generally be used first before a second column of the same type is used. Often a change in the buffer pH will alter how proteins interact with a given column (excluding size exclusion), and can assist in the final stages of separation. Of course, certain pH values, buffers, columns, etc. may cause instability of the protein of interest and must

be avoided once determined. The purification may need to be repeated and purification steps added, removed, or rearranged in order to optimize the purification and maximize the chances of success. A purification table such as in Table 1 is useful for tracking the effectiveness of each purification step. Degree of purification is routinely quantified by enzyme activities, but can alternatively be tracked using the amount of metal. Note, however, that there will likely be multiple proteins present that contain the same element, particularly with less pure samples and the more abundant metals. Hence, if in a chromatography step the target protein is removed from a more abundant protein that contains the same metal, there will be a significant decrease in the amount of the metal present in the target sample.

ANTICIPATED RESULTS

Using Basic Protocol 1 with careful contaminant-free sample preparation and a properly tuned instrument, multiple metals can be measured to a high degree of accuracy with low standard deviation between replicates. The detection limit of specific metals is element specific and depends on the instrument that is used.

Basic Protocol 2 results in the creation of a series of wash samples that contain decreasing amounts of unbound metals and other small molecules, and two S100 samples: the original unwashed S100 and the final 256-fold buffer-exchanged S100 (S100w). ICP-MS analysis is used to determine the concentration of each metal in each sample. By analyzing the distribution of each metal in the original S100 and S100w, as well as all the intermediate washes, one can make a determination of the kinds of metal-macromolecule interactions within the soluble fraction of any organism for any growth condition. Fig 8 depicts the general profiles that define the types of interactions that can be identified.

In Figure 8A, all the metal from the S100 is found primarily in the flow-through, with a negligible amount in the S100w, clearly indicative of unbound or very weakly-bound metal that is easily released. In Figure 8B, all the metal from the S100 is retained in the S100w, which indicates that all the metal in the cell is present as a tightly-bound form in a macromolecule (> 3 kDa). In Figure 8C approximately one third of the metal appears to be tightly-bound and is retained in the S100w, whereas the remainder of the metal is unbound or very weakly-bound and is released into the FT and washes. These data indicate both unbound or weakly-bound metals as well as tight metal-macromolecule interactions. The profiles of the tightly-bound metal-macromolecules, the values found in the S100w, can also be used to determine the relative abundance of the different metals within the cytoplasm.

Basic Protocol 3 has only been carried out a handful of times, as published, and all of the known examples were performed using the same starting sample, soluble cytoplasmic extract of *P. furiosus* (5). Thus, the success of this procedure in other systems may be variable. That said, there are undoubtedly unknown metalloproteins in every organism. It can be expected that at least five chromatography steps will be required to purify the protein of interest to homogeneity and several iterations and refinements of the purification may be required.

TIME CONSIDERATIONS

Protocol 1 time requirements vary greatly with the number of samples analyzed. With the use of an autosampler, sample introduction and data acquisition requires approximately 5-10 minutes per sample. A chromatographic separation such as Basic Protocol 3 typically yields 50-100 fractions per column which, when analyzed in duplicate or triplicate requires 8-50 hours of total run time. Due to space limitations of the autosampler deck and the need for consistant

argon flow rate, chromatographic fractions may need to be analyzed in multiple runs. Plasticware preparation requires 4-5 days, but many tubes can be processed at one time and stored for later use. Sample preparation takes a minimum of 2-3 hours, with additional time depending on the number of samples. Instrument tuning and maintenance time varies by instrument and with the types of samples introduced.

Protocol 2 can be completed in a week. This includes two days to plan and prepare for the experiment, one to two days to make the extract and do the buffer washes (depending on the number of samples) and an additional day or two to carry out the protein estimations and analyze the ICP-MS metal data, as detailed below.

Day 1: Plan experiment. Collect all required components and put glassware into acid-bath for overnight soaking.

Day 2: Rinse out and dry the acid-washed glassware, label everything and place into the anaerobic chamber overnight to allow all diffused oxygen to be scrubbed out by the catalyst.

Day 3/4: Make anaerobic lysis/wash buffer. If possible, break the cells, prepare the 100,000 x *g* supernatant and go through the wash protocol on the same day. If you cannot finish in one day, stop after making the S100 and store the sample(s) overnight at 4°C and start the wash protocol the next morning.

Day 4/5: Perform protein estimation on the samples to ensure you have expected results (see Step 12 in protocol) and give samples for ICP-MS.

Day 5/6: ICP-MS and metal/protein data analysis.

Basic Protocol 3 can be completed in a minimum of about four weeks. As mentioned above, for discovery of a new protein or application of the protocol to a different organism or sample source, the entire protocol will need to be repeated. The first column requires about five days to run and analyze all data due to its large size, but subsequent columns take about three days each. With the smaller columns the first day is preparatory and involves reviewing data from the previous column and selecting fractions to pool as well as preparing and degassing buffers, stoppering and labeling the vials or bottles for fraction collection and setting up the column. The second day is occupied with running the column, sampling the fractions and submitting samples for ICP-MS. The third day allows time for running SDS-PAGE gels, protein estimation of the fractions, and generation and analysis of ICP-MS data. ICP-MS analysis can easily become a bottleneck in this process and may extend the overall time required to complete the protocol.

ACKNOWLEDGEMENTS

The research reported herein was part of the ENIGMA project and was funded by the Office of Science, Office of Biological and Environmental Research, of the U. S. Department of Energy under Contract No. DE-AC02-05CH11231.

REFERENCES

1. **Agatemor, C., and D. Beauchemin.** 2011. Matrix effects in inductively coupled plasma mass spectrometry: a review. Anal Chim Acta **706:**66-83.

2. **Bertini, I., H. B. Gray, E. I. Stiefel, and J. S. Valentine (ed.).** 2007. Biological Inorganic Chemistry: Structure and Reactivity. University Science Books, Sausalito.

3. **Broadhurst, A. V.** 2001. Hydroxylapatite Chromatography, Current Protocols in Protein Science. John Wiley & Sons, Inc.

4. **Brunhold, T. C.** 2007. Synthetic iron-oxo "diamond core" mimics structure of key intermediate in methane monooxygenase catalytic cycle. Proc. Natl. Acad. Sci. U. S. A. **104:**20641-20642.

5. **Cvetkovic, A., A. L. Menon, M. P. Thorgersen, J. W. Scott, F. L. Poole, II, 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. W. Adams.** 2010. Microbial metalloproteomes are largely uncharacterized. Nature (London, U. K.) **466:**779-782.

6. **Dunn, B. M.** 2001. Conventional Chromatographic Separations, Current Protocols in Protein Science. John Wiley & Sons, Inc.

7. **Frey, P. A., A. D. Hegeman, and F. J. Ruzicka.** 2008. The Radical SAM Superfamily. Crit. Rev. Biochem. Mol. Biol. **43:**63-88.

8. **Gallagher, S. R.** 2008. SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE), Current Protocols Essential Laboratory Techniques. John Wiley & Sons, Inc.

9. **Guskov, A., A. Gabdulkhakov, M. Broser, C. Glockner, J. Hellmich, J. Kern, J. Frank, F. Muh, W. Saenger, and A. Zouni.** 2010. Recent progress in the crystallographic studies of photosystem II. Chemphyschem **11:**1160-71.

10. **Ha, Y., O. G. Tsay, and D. G. Churchill.** 2011. A tutorial and mini-review of the ICP-MS technique for determinations of transition metal ion and main group element concentration in the neurodegenerative and brain sciences. Monatsh. Chem. **142:**385-398.

11. **Irvine, G. B.** 2001. Determination of Molecular Size by Size-Exclusion Chromatography (Gel Filtration), Current Protocols in Cell Biology. John Wiley & Sons, Inc.

12. **Kennedy, R. M.** 2001. Hydrophobic-Interaction Chromatography, Current Protocols in Protein Science. John Wiley & Sons, Inc.

13. **Lubitz, W., E. Reijerse, and M. van Gastel.** 2007. [NiFe] and [FeFe] Hydrogenases Studied by Advanced Magnetic Resonance Techniques. Chem. Rev. (Washington, DC, U. S.) **107:**4331-4365.

14. **Matthews, R. G.** 2001. Cobalamin-dependent methyltransferases. Acc. Chem. Res. **34:**681-689.

15. **Paquete, C. M., and R. O. Louro.** 2010. Molecular details of multielectron transfer: the case of multiheme cytochromes from metal respiring organisms. Dalton Trans. **39:**4259-4266.

16. **Ragsdale, S. W.** 2009. Nickel-based enzyme systems. J. Biol. Chem. **284:**18571-18575.

17. **Schwarz, G., R. R. Mendel, and M. W. Ribbe.** 2009. Molybdenum cofactors, enzymes and pathways. Nature (London, U. K.) **460:**839-847.

18. **Sevcenco, A.-M., L. E. Bevers, M. W. H. Pinkse, G. C. Krijger, H. T. Wolterbeek, P. D. E. M. Verhaert, W. R. Hagen, and P.-L. Hagedoorn.** 2010. Molybdenum incorporation in tungsten aldehyde oxidoreductase enzymes from Pyrococcus furiosus. J. Bacteriol. **192:**4143-4152.

19. **Williams, A., and V. Frasca.** 2001. Ion-Exchange Chromatography, Current Protocols in Protein Science. John Wiley & Sons, Inc.

Table B.1. A purification table showing the purification and yield of the molybdenum in PF1972 over five chromatography steps. Percent yield is calculated by the amount of molybdenum in the pool from the current column divided by the amount of molybdenum observed in the pool from the first column (Mo_{Cn} / Mo_{C1}). The fold purification is calculated by the molybdenum to protein ratio for the current column divided by the molybdenum to protein ratio for the first column ($Mo:P_{Cn}$ / $Mo:P_{C1}$).

Column Level	Column Type	Pool volume (mL)	Total protein (mg)	Total Mo (nmol)	Metal to protein ratio (nmol/mg)	Yield (%)	Fold purification
C1	DEAE	488	55.2	11.5	0.21		1.0
C2	SX75 26/60	13.8	1.64	2.30	1.40	20.1	6.7
C3	HAP (8mL)	6.6	0.05	2.53	51.1	22.0	245.3
C4	Phenyl sepharose	9.7	0.03	0.20	7.61	1.75	36.6
C5	Mono Q	1.0	0.005	0.12	140.7	1.1	118.5

Figure B.1. Flowchart showing the preparation of the cytoplasmic (soluble) fraction and the series of buffer exchange steps that the soluble fraction undergoes to generate the 256-fold washed cytoplasmic extract. The fractions generated at each step are analyzed by ICP-MS.

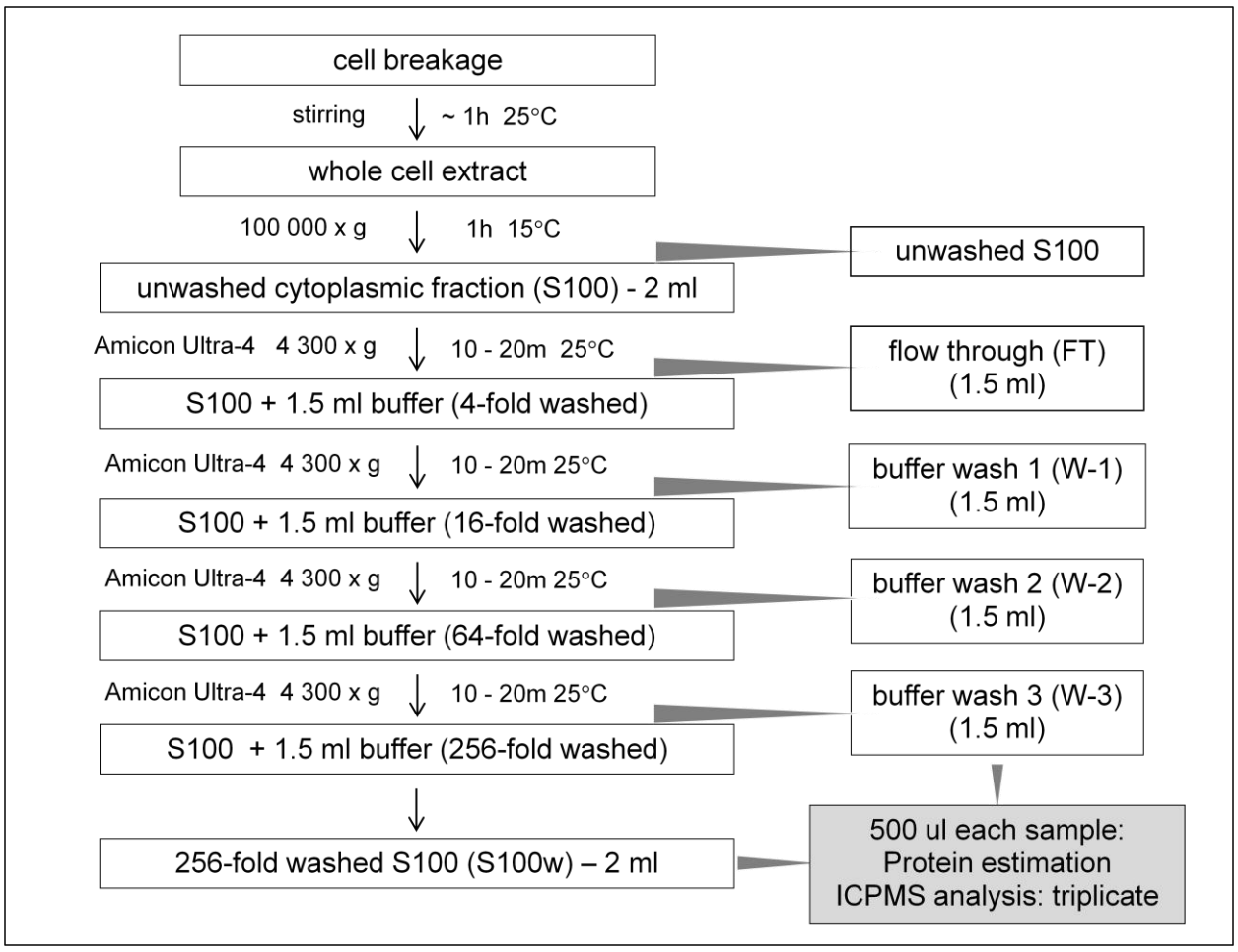

cell breakage
stirring
~ 1h 25°C
whole cell extract
100 000 x g
1h 15°C
unwashed S100
unwashed cytoplasmic fraction (S100) - 2 ml
Amicon Ultra-4 4 300 x g
10 - 20m 25°C
flow through (FT)
(1.5 ml)
S100 + 1.5 ml buffer (4-fold washed)
Amicon Ultra-4 4 300 x g
10 - 20m 25°C
buffer wash 1 (W-1)
(1.5 ml)
S100 + 1.5 ml buffer (16-fold washed)
Amicon Ultra-4 4 300 x g
10 - 20m 25°C
buffer wash 2 (W-2)
(1.5 ml)
S100 + 1.5 ml buffer (64-fold washed)
Amicon Ultra-4 4 300 x g
10 - 20m 25°C
buffer wash 3 (W-3)
(1.5 ml)
S100 + 1.5 ml buffer (256-fold washed)
256-fold washed S100 (S100w) – 2 ml
500 ul each sample:
Protein estimation
ICPMS analysis: triplicate

Figure B.2. Anaerobic chromatography set-up showing anaerobic buffer introduction, anaerobic sample collection, and the direction of buffer and argon flow.

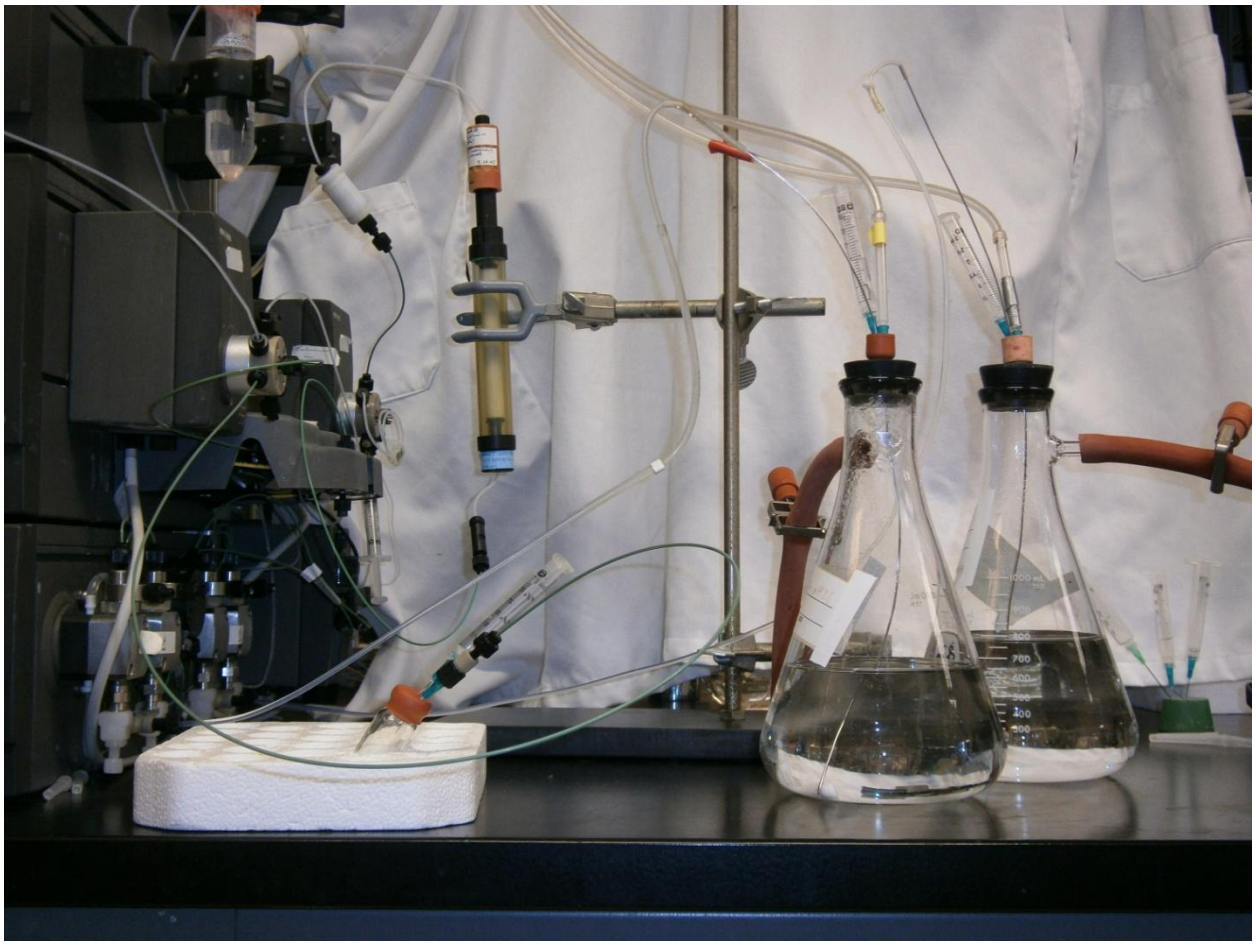


Figure B.3 Elution of soluble Pf proteins from a DEAE anion exchange column showing molybdenum (-x-) and protein (··o··) concentrations. Box shows fractions pooled for PF1972 purification (adapted from Cvetkovic *et al.* 2010).

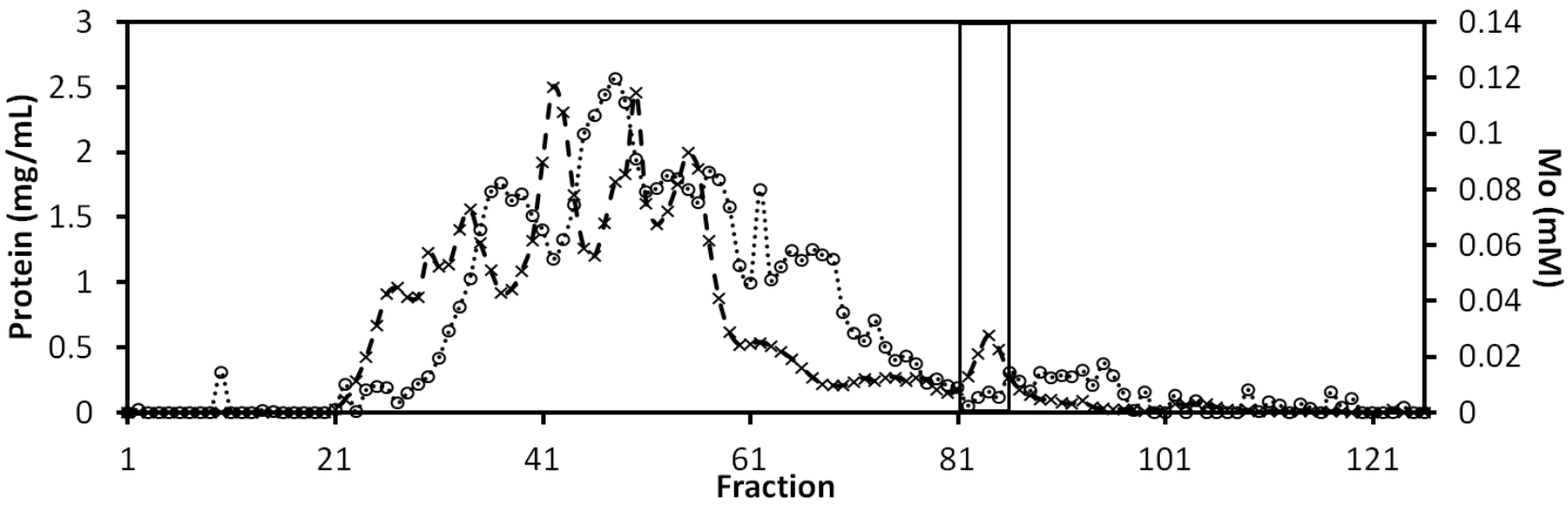

Protein (mg/mL)
Mo (mM)
Fraction
3
2.5
2
1.5
1
0.5
0
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
1
21
41
61
81
101
121

Figure B.4. Elution from second level size exclusion column of the fraction pool from Figure 3 showing molybdenum (-x-) and protein (··o··) concentrations. Box shows fractions pooled for PF1972 purification, (adapted from Cvetkovic *et al.* 2010).

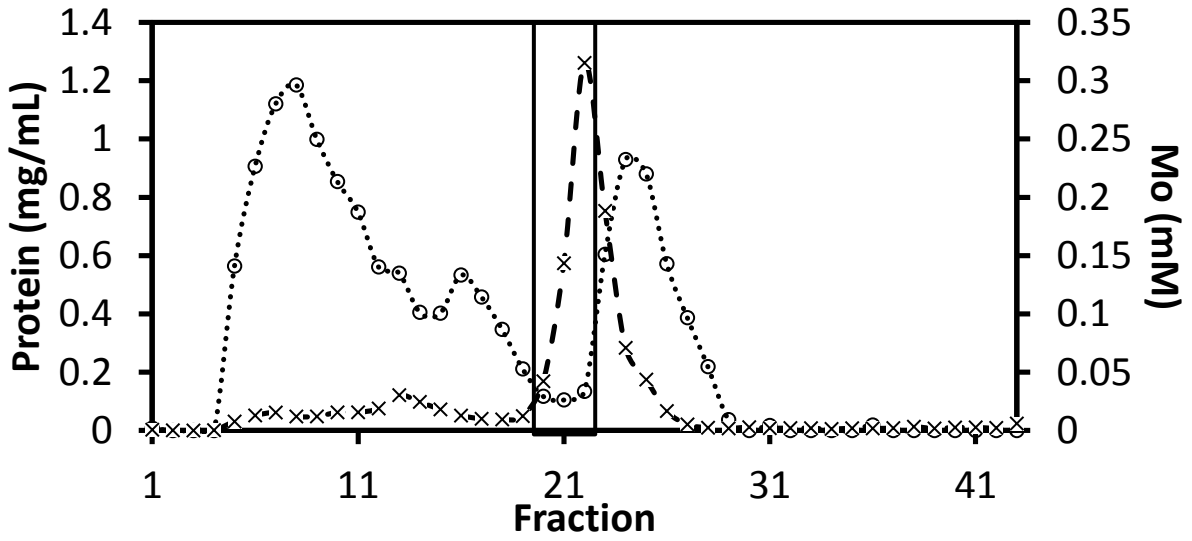

1.4
1.2
1
0.8
0.6
0.4
0.2
0
Protein (mg/mL)
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Mo (mM)
1
11
21
31
41
Fraction

Figure B.5 Elution from a third level hydroxyapatite column of fraction pool from Figure 4 showing molybdenum (-x-) and protein (··o··) concentrations. Box shows fractions pooled for PF1972 purification (adapted from Cvetkovic *et al.* 2010).

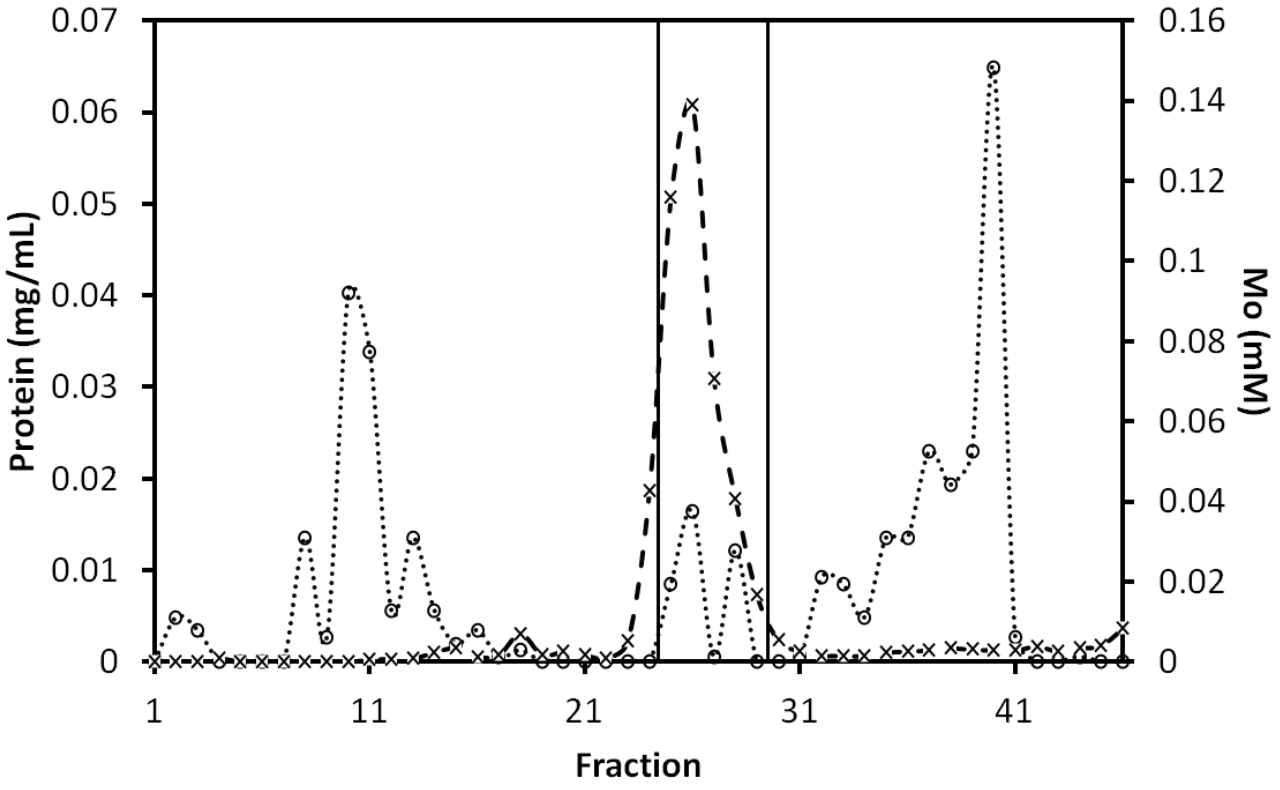

0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
Protein (mg/mL)
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
Mo (mM)
1
11
21
31
41
Fraction

Figure B.6 Elution from a fourth level phenyl sepharose column of fraction pool from Figure 5 showing molybdenum concentrations (-x-). Box shows fractions pooled for PF1972 purification (adapted from Cvetkovic *et al.* 2010).

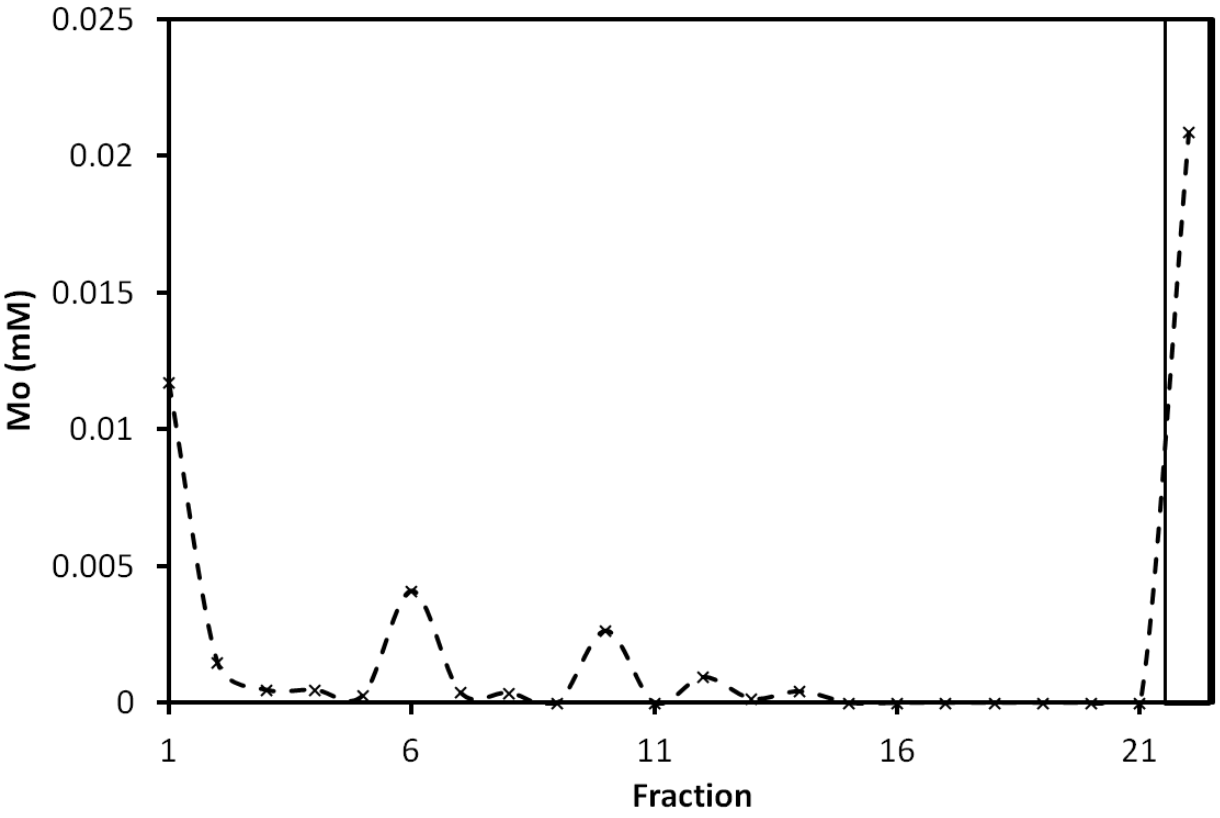
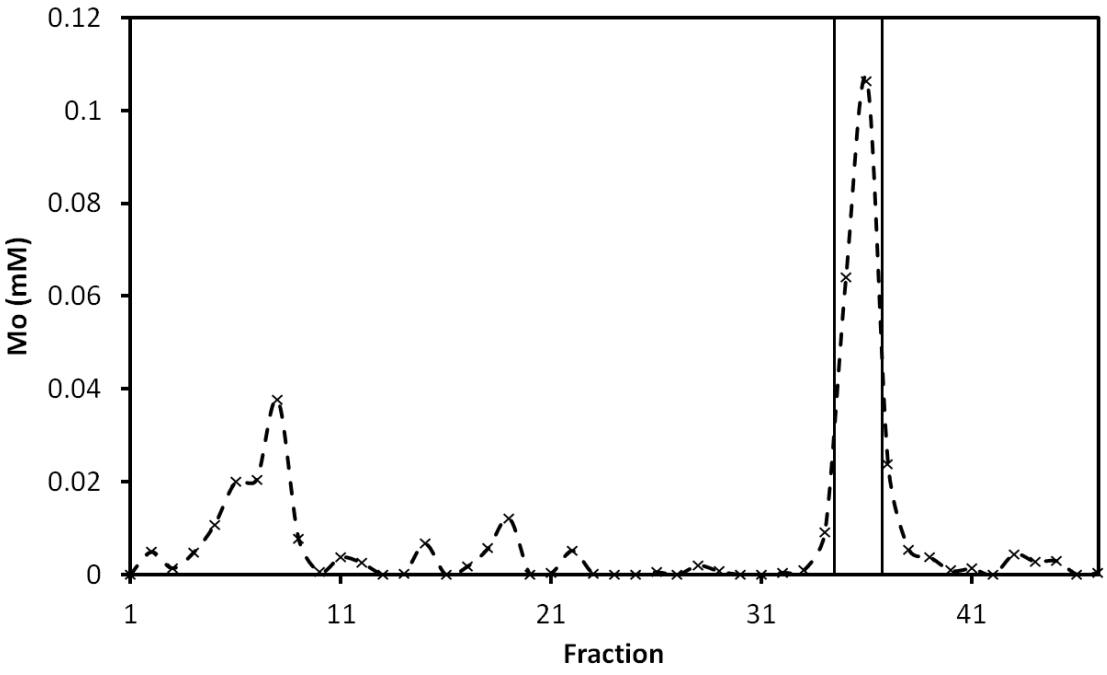

0.025
0.02
0.015
0.01
0.005
0
Mo (mM)
1
6
11
16
21
Fraction

Figure B.7 a) Elution from a fifth level monoQ column of fraction pool from Figure 6 showing molybdenum concentrations (-x-). Box shows purified PF1972. b) SDS-PAGE gel of purified PF1972 (adapted from Cvetkovic *et al.* 2010)

A)


0.12
0.1
0.08
0.06
0.04
0.02
0
Mo (mM)
1
11
21
31
41
Fraction

B)

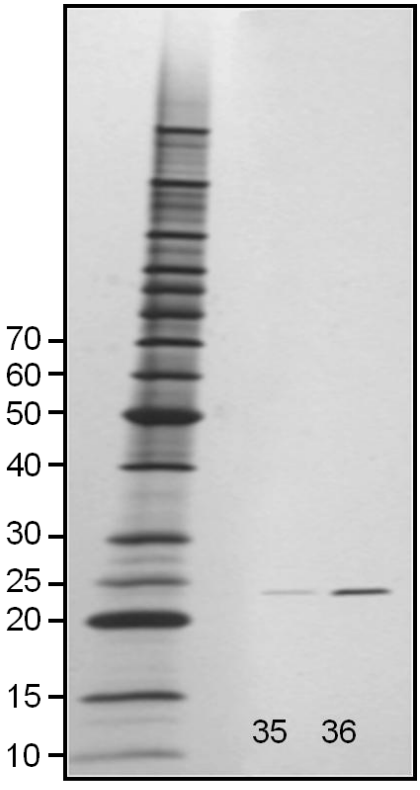
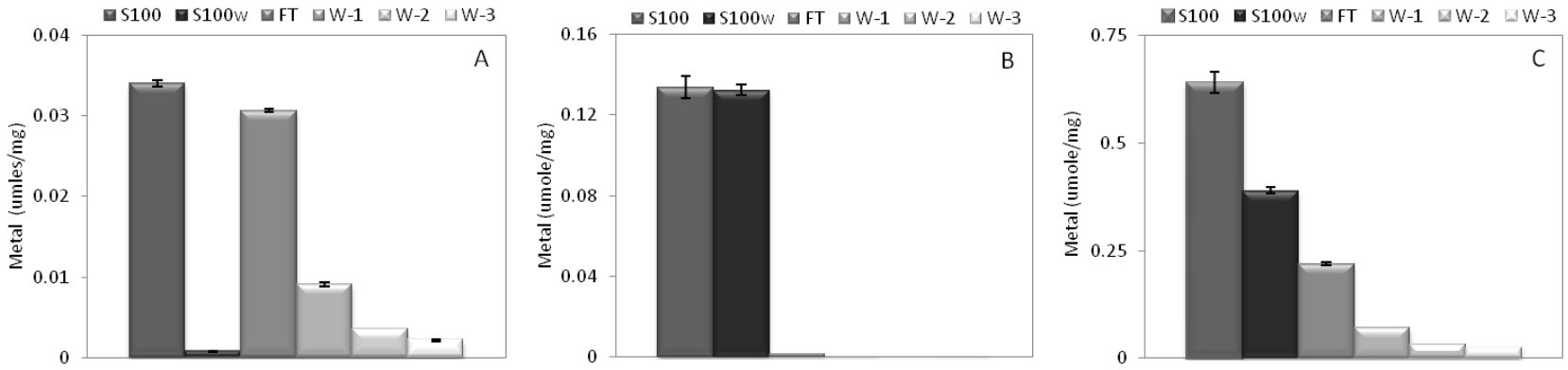

70
60
50
40
30
25
20
15
10
35
36

Figure B.8. The susceptibility of metals to removal from the cytoplasmic fraction by filtration (3 kDa cut-off). In each plot, the bars represent (from left to right) the amount of metal per mg protein in the cytoplasm before (S100) and after washing (S100w), the metal in the flow-through (FT) and in the three subsequent wash steps (W-1, W-2 and W-3). The three bar graphs show representative examples of metal-macromolecule interactions. A: unbound metal/weak interactions; B: tight interactions and C: mixed (a mixture of both types of interactions).


S100
S100w
FT
W-1
W-2
W-3
A
Metal (umles/mg)
0.04
0.03
0.02
0.01
0
B
Metal (umole/mg)
0.16
0.12
0.08
0.04
0
C
Metal (umole/mg)
0.75
0.5
0.25
0

APPENDIX C

MICROBIAL METALLOPROTEOMES ARE LARGELY UNCHARACTERIZED[6]

[6] **Cvetkovic, A.*, A. L. Menon*, M. P. Thorgersen*, J. W. Scott, F. L. Poole, II, 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. W. Adams.** 2010. Microbial metalloproteomes are largely uncharacterized. Nature (London, U. K.) **466:**779-782.
*Contributed equally to this work

ABSTRACT

Metal ion cofactors afford proteins virtually unlimited catalytic potential, enable electron transfer reactions and have a great impact on protein stability(13, 22). Consequently, metalloproteins have key roles in most biological processes, including respiration (iron and copper), photosynthesis (manganese) and drug metabolism (iron). Yet, predicting from genome sequence the numbers and types of metal an organism assimilates from its environment or uses in its metalloproteome is currently impossible because metal coordination sites are diverse and poorly recognized(17, 22, 32). We present here a robust, metal-based approach to determine all metals an organism assimilates and identify its metalloproteins on a genome-wide scale. This shifts the focus from classical protein-based purification to metal-based identification and purification by liquid chromatography, high-throughput tandem mass spectrometry (HT-MS/MS) and inductively coupled plasma mass spectrometry (ICP-MS) to characterize cytoplasmic metalloproteins from an exemplary microorganism. Of 343 metal peaks in chromatography fractions, 158 did not match any predicted metalloprotein. Unassigned peaks included metals known to be used (cobalt, iron, nickel, tungsten and zinc; 83 peaks) plus metals the organism was not thought to assimilate (lead, manganese, molybdenum, uranium and vanadium; 75 peaks). Purification of eight of 158 unexpected metal peaks yielded four novel nickel- and molybdenum-containing proteins, whereas four purified proteins contained sub-stoichiometric amounts of misincorporated lead and uranium. Analyses of two additional microorganisms (*Escherichia coli* and *Sulfolobus solfataricus*) revealed species-specific assimilation of yet more unexpected metals. Metalloproteomes are therefore much more extensive and diverse than previously recognized, and promise to provide key insights for cell biology, microbial growth and toxicity mechanisms.

INTRODUCTION

Once revealed, a metal cofactor adds new dimensions to understanding protein structure and function; yet, the presence of metal is often unsuspected until the protein is analysed (13, 17, 22). For example, unexpected zinc and iron–sulphur sites gave fundamental insights into DNA repair proteins relevant to human cancers(11). Unfortunately, the small fraction of biochemically characterized proteins and limitations of metalloprotein bioinformatics (4, 9, 13, 17, 34, 36) make it impossible to predict metals used by organisms and to define any metalloproteome. Previous metal-based studies examined individual purified proteins, recombinant proteins, biological fluids (such as blood or urine) or involved limited metals (5, 19, 28, 31, 34). Yet, native biomass is likely essential as metalloproteins from recombinant sources may have incorrect or no metal at all(10, 15).

Herein we present combined technologies that reveal assimilated metals and metalloproteins from biomass of the prototypical microbe *Pyrococcus furiosus* (12) (Supplementary Fig. 1). Whereas proteins containing five transition metals, cobalt (Co), iron (Fe), nickel (Ni), tungsten (W) and zinc (Zn), have been purified from *P. furiosus* (Supplementary Table 1) previously, surprisingly 21 of 53 metals analysed by ICP-MS (28) were detected in the cytoplasmic extract, including lead (Pb), titanium (Ti) and uranium (U) (Supplementary Tables2 and 3). *P. furiosus* specifically assimilated these 21 from 44 metals in the growth medium (Supplementary Table 4). This had seven added metals, the remaining coming from added organic components. Excepting chromium (Cr), ruthenium (Ru) and strontium (Sr), 18 metals were in macromolecular complexes (≥5 kDa) rather than free ions (Supplementary Fig. 2). Cells grown with added Pb, U, Ru, rhodium (Rh, each 50 nM) and Cr

(200 nM) contained a more than tenfold increase in the intracellular concentrations of tightly-bound Pb and U, but not Rh, Ru and Cr (Supplementary Fig. 3), indicating specific uptake of metals that are available to *P. furiosus* in its marine environment(3).

To investigate if uptake of unanticipated metals involved biological functions or inadvertent assimilation, we examined stable cytoplasmic metalloproteins that retained metals after an anion exchange separation (chromatography 1 (C1); Supplementary Fig. 4)(21). This unambiguously identified 10 metals as multiple peaks in 126 C1 chromatography fractions: (1) molybdenum (Mo), manganese (Mn) and vanadium (V), not previously known in *P. furiosus*, (2) U and Pb (not found previously in any organism except in detoxification proteins), and (3) known metals (Co, Fe, Ni, W and Zn). Other cytoplasmic metals had no distinct peaks in C1 fractions. The most abundant were Fe and Zn (97% of the total), with less W and Ni (<2.5%) and even less Co, Mo, Mn, Pb, V and U (<0.5%; Fig. 1a, Supplementary Tables 5 and 6).

To further explore the *P. furiosus* metalloproteome, we separated C1 fractions by second level chromatography (C2; Supplementary Fig. 5). ICP-MS of 790 fractions obtained from fifteen C2 columns revealed 343 distinct metal peaks (Fig. 2, Supplementary Tables 7 and 8). HT-MS/MS(21) identified 770 proteins or ~60% of the cytoplasmic proteins(24). Given the difficulties of predicting metalloproteins (4, 9, 13, 17, 36), we searched the Integrated Resource of Protein Domains and Functional Sites (InterPro) (14) annotation of the *P. furiosus* genome (Supplementary Table 9) to assign metal peaks to proteins. This InterPro-Metal (IPM) analysis identified domains even remotely related to those that bind a metal. However, only 185 of the 343 metal peaks detected in the C2 fractions contained an IPM-predicted protein for the identified metal. The remaining 158 unassigned metal peaks therefore represent metalloproteins containing unknown metal-binding domains (Fig. 1c, Supplementary Table 8). Consequently, *P.*

furiosus assimilates more metals than expected, and even metals it is known to use (Co, Fe, Ni, W and Zn) give rise to numerous unassigned peaks (>80).

To test the feasibility of assigning proteins to the 158 metal peaks without IPM hits, we selected eight peaks (two Mo, Ni, Pb and U; Supplementary Table 8) for multistep chromatography purification to obtain a homogeneous protein containing a near stoichiometric amount of the metal, analogous to the traditional purification of an enzymatic activity (Supplementary Table 10). Using 300 g of *P. furiosus* biomass, one of the 29 unassigned Ni peaks after five chromatography steps yielded a pure Ni-containing protein (600 µg) identified by matrix-assisted laser desorption/ionization (MALDI)-MS as PF0056 (Supplementary Fig. 6). This cupin/putative sugar-binding protein (14 kDa) had no IPM hit for Ni yet contained 0.47 ± 0.05 Ni and 0.50 ± 0.08 Zn atoms per mole (but no other > 0.1 atoms mol^{-1}). The PF0056 metal ions are predicted to be coordinated by three His and one Glu residue, from a homologue structure (PDB 1VJ2, Table 1). Cupins are among the most functionally diverse protein superfamilies (2); PF0056 is the first native Ni-containing member of this family to be purified.

Two of 18 Mo peaks lacking IPM-predicted molybdoproteins were also purified (Supplementary Figs 7 and 8). After five chromatography steps, one Mo peak yielded a homogeneous protein (2.5 µg) identified by MALDI-MS as PF1972 (27.6 kDa). It contained Mo (0.77 ± 0.43 atoms mol^{-1}) and also iron with a Fe:Mo ratio of 4:1 (Supplementary Fig. 7) and is predicted to be a [4Fe–4S] cluster-containing activase for anaerobic ribonucleotide reductase (PF1971)(20). Activases are widespread in anaerobes, but those in hyperthermophiles, like *P. furiosus*, contain four conserved Cys residues besides three expected conserved Cys residues coordinating the [4Fe–4S] cluster. Expression of PF1972 is upregulated at suboptimal growth temperatures, indicating a role for Mo in DNA synthesis under these conditions (35). The second

Mo peak co-purified with two proteins that partially separated after six chromatography steps: a known tungstoprotein (PF0464; Supplementary Table 1) lacking Mo, and PF1587 with unique peptides detected by HT-MS/MS matching the Mo peak (Supplementary Fig. 8). PF1587 is a 32.9 kDa conserved hypothetical protein with archaeal and bacterial homologues that contain five conserved cysteines. If the cysteine residues in PF1587 and PF1972 directly coordinate Mo, this would be unprecedented for molybdoenzymes, wherein Mo is bound by S atoms of an organic pterin cofactor (30). PF1972 and PF1587 are the first Mo-proteins purified from *P. furiosus*, an organism not previously known to assimilate this metal.

The second unassigned Ni peak purified originated from PF0086 on the basis of native biomass (21) and recombinant protein data. After three chromatography steps (Fig. 2, Supplementary Fig. 9), only PF0086 had a profile of unique peptides that matched the Ni peak (Supplementary Fig. 9 and Table 11). Annotated as alanyl-tRNA editing hydrolase, PF0086 is predicted to contain Zn (33). but Zn was undetected in PF0086 fractions. To confirm PF0086 is a bona fide Ni-protein, the corresponding gene was expressed in *E. coli* grown in a Ni-supplemented medium (200 μM). The purified protein contained 0.86 ± 0.20 Ni atoms mol^{-1}. Interestingly, when PF0086 was expressed in *E. coli* grown in a Zn- or Co-supplemented medium (200 μM), its predominant metal was Zn or Co, respectively (Supplementary Fig. 10). *E. coli* evidently inserts the most abundant metal (Ni, Co or Zn), whereas *P. furiosus* specifically inserts Ni into PF0086, despite a Zn concentration ~50-fold greater than Ni in its growth medium (Fig. 1b). On the basis of a homologue structure (PDB 2E1B), the Ni in PF0086 is likely coordinated by three His and one Cys residue (Table 1). PF0086 is another new type of Ni-containing enzyme and with PF0056, increases the number known in biology from eight to ten

(26). Such Ni-enzyme discoveries can enhance understanding of catalysis and biology as seen for Ni versus other metal-containing superoxide dismutases (23).

In contrast to Ni and Mo, purifications of U and Pb peaks, even from cells grown with more than tenfold higher U and Pb concentrations, yielded homogeneous proteins with only trace amounts of these metals (Table 1; Supplementary Table 12). For example, purification of one of 34 Pb peaks through six chromatography steps (Supplementary Fig. 11) yielded a single protein identified as PF1343, a known metalloprotease (39.4 kDa), containing 0.65 ± 0.09 Zn atoms mol^{-1} but only 0.010 ± 0.002 Pb atoms mol^{-1} (Table 1). Similarly, after five chromatography steps, another Pb peak yielded homogeneous PF0257, a pyrophosphatase (20.9kDa; Supplementary Fig. 12). This is a new iron protein (1.40 ± 0.02 Fe atoms mol^{-1}) with low amounts of Pb (0.007 ± 0.001 Pb atoms mol^{-1}).

Similarly, after three chromatography steps, one U peak was associated with known iron-protein ferritin (PF0742, 20.3 kDa; Table 1, Supplementary Fig. 13) containing 1.20 ± 0.11 Fe atoms mol^{-1} but only 0.010 ± 0.001 U atoms mol^{-1} (and also 0.010 ± 0.001 Pb atoms mol^{-1}). A second U peak copurified through six steps with the glycolytic Mg^{2+}-dependent enzyme enolase (PF0215, 46.8 kDa) but contained only 0.00010 ± 0.00004 U atoms mol^{-1} (Supplementary Fig. 14). Although other U and Pb peaks may represent bona fide U- and Pb-proteins, the four analysed seem to have misincorporated U and Pb that dissociate over multiple chromatography steps. However, identification of proteins susceptible to such metal misincorporation has implications in elucidating mechanisms of metal toxicity in both prokaryotes and eukaryotes. Our approach can identify proteins containing any of 53 metals using any organism's biomass without requiring radiolabels.

To test further if this approach is generally applicable, we fractionated cytoplasmic extracts of *Escherichia coli* and *Sulfolobus solfataricus* (37). Although their growth media contained the same 44 metals as the *P. furiosus* medium (Supplementary Table 4), there were substantial differences in the metals assimilated (Supplementary Figs 15 and 16; Table 13). Their C1 fractions also contained distinct peaks of Co, Fe, Mo, Mn, V, Zn and Pb, but *E. coli* fractions uniquely contained cadmium (Cd) and arsenic (As), whereas tin (Sn) and antimony (Sb) were found only in *S. solfataricus*. *E. coli* fractions also contained U and Ni but those of *S. solfataricus* did not. Which assimilated metals are biologically functional can be ascertained by the methods described herein. Such organism-specific assimilation likely reflects natural environments. *P. furiosus* is a marine anaerobe, *S. solfataricus* is a freshwater aerobic acidophile, and *E. coli* is a facultative anaerobe inhabiting the human gut. In general metal availability could significantly alter the metalloproteome and consequently microbial physiology, which raises the issue of whether laboratory media satisfy organisms' metal requirements. This has an impact on efforts to grow new microbes and communities, which are often challenging or impossible.

Overall, we find that much of microbial metalloproteomes remain uncharacterized. Remarkably, even with metals *P. furiosus* was known to assimilate, half of the observed peaks were unassigned (Fig. 1c). Given the major roles that metals have in protein function, native metalloproteomes must be characterized to complement recombinant efforts including structural genomics. These results validate our metal-based, non-radiolabel approach to determine metals an organism assimilates and identify new metal-containing proteins with uncharacterized metal-binding domains. These encompass both known protein families and uncharacterized portions of genomes comprised of conserved/hypothetical proteins (Table 1). Furthermore, this technique can identify proteins with misincorporated metals, providing insight into metal toxicity

mechanisms in both organisms and tissues(7, 18). The power and flexibility of this metal-based approach makes it a valuable tool in unlocking a more complete understanding of the far-reaching roles of metals in biology.

METHODS SUMMARY

Pyrococcus furiosus (DSM 3638^T) was grown at 90 C using maltose and peptides as the carbon source and cells were collected at late-exponential phase (21). The cytoplasmic extract was prepared and fractionated by two chromatography steps (C1 and C2) and proteins were identified in the chromatography fractions by HT-MS/MS as described elsewhere (21). Metal concentrations were measured in the uninoculated growth medium, in the cytoplasmic extract, and in C1 and C2 column fractions using a quadrupole-based ICP-MS equipped with a MicroMist Nebulizer operated under Ar with and without He as collision gas. Selected metal peaks in the C2 chromatography fractions were purified further individually by following the metal through multiple chromatography steps until a single protein band was obtained after analysis by sodium dodecyl sulphate electrophoresis. Metal stoichiometry in the purified proteins is based on the molecular mass calculated from the gene sequence and a colorimetric estimate of protein concentration(21). Heterologous expression of PF0086 was induced by isopropyl β-D-1-thiogalactopyranoside in *E. coli* BL21 (DE3) grown aerobically in a rich medium. The recombinant protein was purified by heat treatment (80 C for 15 min) followed by multistep column chromatography. For metal analyses of their C1 chromatography fractions, *E. coli* was grown aerobically at 37 °C in a rich medium and *S. solfataricus* P2 was grown aerobically at 80°C in a medium containing sucrose and peptides at pH 3.0 (37). Cytoplasmic extracts of each

organism were subjected to anion exchange chromatography and metal analysis using the procedures devised for *P.* Furiosus (21). See Methods for details.

SUPPLEMENTARY INFORMATION

Supplementary information is linked to the online version of the paper at www.nature.com/nature.

ACKNOWLEDGEMENTS

This research is part of the MAGGIE (Molecular Assemblies, Genes and Genomes Integrated Efficiently) project supported by Department of Energy grant (DE-FG0207ER64326). We thank S. Hammond, L. Wells, R. Hopkins and D. Phillips for help with in-gel MS analyses.

AUTHOR CONTRIBUTIONS

A.C., A.L.M., M.P.T. and J.W.S. grew and fractionated *P. furiosus*; A.L.M. carried out cytoplasmic washes; A.L.M. and S.M.Y. grew and fractionated *S. solfataricus*; A.L.M. and M.P.T. grew and fractionated *E. coli*; A.C. and S.S. performed ICP-MS analyses; S.A.T., E.K., J.V.A. and G.S. performed HT-MS/MS analyses; A.L.M. purified PF0056; J.W.S. purified PF1972 and PF0086; M.P.T. and B.J.V. purified PF0742; M.T.P. purified PF1587, PF0215, PF1343 and PF0257; W.A.L., J.L.P. and F.L.P. carried out metal-protein bioinformatic analyses; A.C., A.L.M., F.E.J., F.L.P., M.P.T. and J.A.T. and M.W.W.A. contributed to experimental design and data analyses, and wrote the paper.

METHODS

Fractionation of *P. furiosus*.

The procedures for the growth of *Pyrococcus furiosus* (DSM 3638^T) at 90°C on a rich medium (RM), the anaerobic preparation of the cytoplasmic extract, and its anaerobic fractionation using one first (C1) and fifteen second (C2) level chromatography columns, have been described elsewhere(21). The column fractionation procedure is summarized in Supplementary Fig. 5. Information on other column steps is described for each protein that was purified. *P. furiosus* was also grown using the RMex and CMMex media. RMex is the RM medium supplemented with lead, uranium, rhodium and ruthenium ($Pb(NO_3)_2$, $UO_2(C_2H_3O_2)_2.2H_2O$, $RhCl_3.3H_2O$ and $RuCl_3.3H_2O$, each at 50 nM) and chromium ($CrCl_3.6H_2O$, 200 nM). CMMex medium is complete maltose medium (1, 29) containing elemental sulphur (31 mM) to which $Pb(NO_3)_2$ and $UO_2(C_2H_3O_2)_2.2H_2O$ were added at 500 nM each. Cells were processed for the C1 fractionation step as described for cells grown in the RM medium (21). For metal analyses and susceptibility of metals to removal by filtration studies, the cytoplasmic extract from cells grown in the RM and RMex media were used. Frozen cells (3 g) were gently lysed by osmotic shock anaerobically under a continuous flow of Ar in 9 ml of 50 mM Tris-HCl (pH 8.0) containing 2 mM sodium dithionite as a reductant and 0.5 μg ml^{-1} DNase I to reduce viscosity. The cell-lysates were centrifuged at 100,000*g* for 1 h at 18°C and the supernatants representing the cytoplasmic fractions were used for the metal analyses.

Protein identification.

The procedures for protein identification in solution using high-throughput tandem mass spectrometry (HT-MS/MS) were described previously (21). Proteins identified by MALDI-MS were first separated using native- or SDS-PAGE gradient gel electrophoresis (4–20% Criterion gels; Bio-Rad). The gel bands of interest were cut out, processed and digested for 16 h at 37°C according to the manufacturer's protocol provided with the recombinant porcine trypsin used for the in-gel protein digest (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 a MTP 384 Massive MADLI target (Bruker Daltonics), along with 1 µl of ProteoMass Peptide & Protein MALDI-MS Calibration Kit standard (Sigma-Aldrich). The target was analysed using a Bruker Daltonics Autoflex MADLI 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.

Metal Analyses

Metals were measured using a quadrupole-based ICP-MS (7500ce, Agilent Technologies) equipped with a MicroMist Nebulizer (Agilent Technologies). This system uses an octupole collision/reaction cell for collision focusing and interference reduction. Sample solutions were introduced into the instrument via a peristaltic pump from an ASX-500 series ICP-MS autosampler (Agilent Technologies) at a flow rate of 0.2 ml min^{-1} into a water-cooled (2 °C) quartz spray chamber. Argon (>99.99% purity) was used as the plasma, auxiliary, nebulizer and makeup gas. The instrument was operated with and without a collision gas. The use of helium (>99.99% purity) as the collision gas effectively removes almost all matrix- and carrier-gas-related interferences (the impact of ArO, a main interference of iron isotope ^{56}Fe, was negligible even when the instrument was run in the collision mode; data not shown). Before analysis, the instrument was stabilized for 30 min and equilibrated for 40 min with a matrix solution identical to that of the sample to be analysed. The operation conditions used to analyse all fractions are summarized in Supplementary Table 2.

Quantification of metals was performed using certified standard reference materials (IV-ICPMS-71A CCS-5 and CMS-2; Inorganic Ventures) as external standards. Standard stock solutions were diluted with high-purity, glass-distilled deionized water obtained from a Corning Mega-Pure System D2 water purifier (Corning) acidified with 2% (v/v) trace metal grade nitric acid (Fisher Scientific). To control the stability of the plasma, drifting and matrix effects, an internal standard IV-ICPMS-71D (10 µg l^{-1} of Li, Sc, Y, In, Tb and Bi; Inorganic Ventures) was automatically added to the samples and to external standards before being added to the nebulizer. Li, Sc and Y were used as internal standards in the presence of collision gas, and Y, In, Tb and Bi were used in the non-collision mode of instrument operation.

Samples were diluted to the desired volume with 2% (v/v) Trace Metal Grade HNO_3 (Fisher Scientific) in acid-washed 15 ml polypropylene tubes (Sarstedt). The acidified samples were vortexed and incubated for at least 1.5 h at 24°C to denature proteins and release metals. The samples were then centrifuged at 2,800*g* for 5 min at 24°C in an Allegra 6R centrifuge (Beckman) immediately before the experimental run. The experimental parameters for ICP-MS (without collision gas) were optimized to maximize sensitivity for the isotopes present in the tuning solution (1 p.p.b.) using a procedure described by the instrument manufacturer (Agilent Technologies). This was followed by optimization of the maximum ion intensities of the isotopes in the tuning solution to minimize isobaric interferences in the presence of the collision gas. The dual ion detector (used in pulse and analogue mode) was calibrated for each of the investigated isotopes. A calibration (0, 0.1, 0.5, 1, 5, 10 and 50 p.p.b.) was performed for each of the metals to be analysed in a specific run and the regression coefficient for each metal was >0.99. Each sample was analysed in duplicate.

To release metals during sample digestion before ICP-MS analysis (8), the most commonly used acid, HNO_3, has the benefits of wide elemental solubility combined with low levels of interference and signal instability (16). Native purified *P. furiosus* rubredoxin (PF1282), a highly stable iron-containing metalloprotein (6), was used to develop the pretreatment procedure for ICP-MS analysis. The effect of HNO_3 concentration (1, 2 and 5%, v/v) and temperature (24 and 80°C) after a 1 h pretreatment on the release of Fe from rubredoxin (at a final concentration of 0.83 μg mL^{-1}) was evaluated by ICP-MS operated in the collision mode. Pretreatment at 24°C in 2% (v/v) HNO_3 for 1 h led to the quantitative release of Fe (data not shown). The same conditions were used to investigate the release of Co, Ni, Mo, W and Zn from experimental samples. Fraction position 48 collected after fractionation of *P. furiosus*

cytosol over a DEAE-Sepharose FF (DEAE-FF, GE Healthcare) column was diluted 50-fold in 2% (v/v) HNO_3 and after the various pretreatments metals were detected by ICP-MS in collision mode. The results confirmed those obtained with rubredoxin (data not shown), namely, a 1 h sample treatment in 2 and 5% (v/v) nitric acid at 24°C for 1 h before ICP-MS analysis gave the same values for the release of the metals in the *P. furiosus* samples. The validity of all of these approaches is illustrated in Supplementary Fig. 1, which shows that there is an excellent correlation between the iron concentrations measured in the C1 chromatography fractions determined by a colorimetric assay and by ICP-MS analysis performed in the reaction and collision modes.

Identification of metalloproteins in *P. furiosus*.

The list of the *P. furiosus* proteins having metal-associated or metal-binding domains was generated by analyzing the *P. furious* genome using the 2007 InterProScan tool (25) and the InterPro (IPR) database (14). Each gene of *P. furiosus* may have several different IPR hits each with a unique IPR identifier, which corresponds to a family, domain or functional site that is accompanied by a well-maintained summary page on the IPR web site (www.ebi.ac.uk/interpro/). All of the information in the summary pages is available in a single downloadable XML file. This summary XML file was searched against a dictionary of specific metal-related words or phrases (i.e. Fe, iron, metal, etc., see Supplementary Table 9) represented by regular expressions using a simple Perl script. These dictionary hits were then manually assessed for accuracy and tabulated. On the basis of this analysis, a list of *P. furiosus* proteins having metal-associated domains was created for each of the metals detected in the C1 fractions (Supplementary Table 7). These lists were then compared with the proteins identified by HT-

MS/MS (21) within each of the metal peaks to determine if any known or predicted protein contained that metal.

Recombinant protein expression.

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 medium, which contains casein hydrolysate, casamino acids and yeast extract (27). 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-l cultures were grown at 37 °C to an A_{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 collected 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*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 48,000g for

20 min. The supernatant was applied to a 5 ml QHP column (GE Healthcare) 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 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. The metal content of the recombinant proteins obtained from *E. coli* grown in various media is shown in Supplementary Fig. 10.

Fractionation of *S. solfataricus* and *E. coli*.

E. coli BW25113 was grown aerobically with shaking at 37°C in 2 l of rich (2×YT) medium and collected in the late log phase yielding 11 g of cell paste (27). All further steps were performed under anaerobic and reducing conditions. The cells were resuspended in 3 CV of 50 mM Tris HCl (pH 8.0) containing 2 mM Na-dithionite (Buffer A) and 0.05 mg ml^{-1} lysozyme and incubated with shaking for 1 h at 25°C. The cell lysate was treated with DNAse I (4 µg ml^{-1}) and incubated an additional 30 min before centrifugation at 47,000*g* for 60 min at 4°C and the supernatant (cytoplasmic fraction) was loaded at 25% in Buffer A onto a 45 ml (5.3×8.5 cm) DEAE-FF column equilibrated in Buffer A. After loading, the column was washed with 5 CV of Buffer A and the proteins were eluted with a 15 CV gradient of 0–500 mM NaCl in Buffer A (60

fractions), followed by a 7 CV gradient of 500–1,000 mM NaCl in Buffer A (5 fractions). The fractions generated were used for ICP-MS analysis (Supplementary Table 13).

S. solfataricus P2 was grown aerobically at 80°C in a 600-l fermenter on a medium containing sucrose and peptides (pH 3.0) and collected in the late log phase yielding ~ 600 g of cell paste(37). The procedure for preparing the anaerobic cell-free extract and for running the first chromatography column were the same as described for *P. furiosus*, except that 100 g of frozen cells were processed. The cytoplasmic fraction was loaded onto a 175 ml (5×9 cm) DEAE-FF column, washed with three CV of Buffer A and proteins were eluted using a 0–250 mM NaCl gradient over 15 CV (60 fractions), followed by a gradient (3 CVs) of 250–1,000 mM NaCl (5 fractions). The fractions generated were used for ICP-MS analysis (Supplementary Table 13).

REFERENCES

1. **Adams, M. W. W., J. F. Holden, A. L. Menon, G. J. Schut, A. M. Grunden, C. Hou, A. M. Hutchins, F. E. Jenney, Jr., C. Kim, K. Ma, G. Pan, R. Roy, R. Sapra, S. V. Story, and M. F. J. M. Verhagen.** 2001. Key Role for Sulfur in Peptide Metabolism and in Regulation of Three Hydrogenases in the Hyperthermophilic Archaeon *Pyrococcus furiosus*. Journal of Bacteriology **183:**716-724.

2. **Agarwal, G., M. Rajavel, B. Gopal, and N. Srinivasan.** 2009. Structure-based phylogeny as a diagnostic for functional characterization of proteins with a cupin fold. PLoS ONE **4:**e5736.

3. **Aiuppa, A., G. Dongarra, G. Capasso, and P. Allard.** 2000. Trace elements in the thermal groundwaters of Vulcano island (Sicily). J. Volc. Geoth. Res. **98:**189-207.

4. **Andreini, C., I. Bertini, G. Cavallaro, G. L. Holliday, and J. M. Thornton.** 2009. Metal-MACiE: a database of metals involved in biological catalysis. Bioinformatics **25:**2088-2089.

5. **Atanassova, A., M. Högbom, and D. B. Zamble.** 2008. High throughput methods for analyzing transition metals in proteins on a microgram scale, p. 319-330. *In* B. Kobe, M. Guss, and T. Huber (ed.), Methods in molecular biology: structural proteomics: high-throughput methods. Humana Press, Tojowa, NJ.

6. **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. W. Adams.** 1991. 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 **30:**10885-10895.

7. **Bressler, J. P., L. Olivi, J. H. Cheong, Y. Kim, A. Maerten, and D. Bannon.** 2007. Metal transporters in intestine and brain: their involvement in metal-associated neurotoxicities. Human & Experimental Toxicology **26:**221-9.

8. **Cai, Y., M. Georgiadis, and J. W. Fourqurean.** 2000. Determination of arsenic in seagrass using inductively coupled plasma mass spectrometry. Spectrochim. Acta B **55:**1411-1422.

9. **Castagnetto, J. M., S. W. Hennessy, V. A. Roberts, E. D. Getzoff, J. A. Tainer, and M. E. Pique.** 2002. MDB: the metalloprotein database and browser at the Scripps Research Institute. Nucl. Acids Res. **30:**379-382.

10. **Chai, S. C., W. L. Wang, and Q. Z. Ye.** 2008. Fe(II) is the native cofactor for *Escherichia coli* methionine aminopeptidase. J Biol Chem **283:**26879-85.

11. **Fan, L., J. O. Fuss, Q. J. Cheng, A. S. Arvai, M. Hammel, V. A. Roberts, P. K. Cooper, and J. A. Tainer.** 2008. XPD xelicase structures and activities: insights into the cancer and aging phenotypes from XPD mutations. Cell **133:**789-800.

12. **Fiala, G., and K. O. Stetter.** 1986. *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100°C. Arch. Microbiol. **145:**56-61.

13. **Gray, H. B., Stiefel, E. I., Valentine J. S. & Bertini, I.** 2006. Biological Inorganic Chemistry: Structure and Reactivity. University Science Books.

14. **Hunter, S., R. Apweiler, T. K. Attwood, A. Bairoch, A. Bateman, D. Binns, P. Bork, U. Das, L. Daugherty, L. Duquenne, R. D. Finn, J. Gough, D. Haft, N. Hulo, D. Kahn, E. Kelly, A. Laugraud, I. Letunic, D. Lonsdale, R. Lopez, M. Madera, J. Maslen, C. McAnulla, J. McDowall, J. Mistry, A. Mitchell, N. Mulder, D. Natale, C. Orengo, A. F. Quinn, J. D. Selengut, C. J. A. Sigrist, M. Thimma, P. D. Thomas, F. Valentin, D. Wilson, C. H. Wu, and C. Yeats.** 2009. InterPro: the integrative protein signature database. Nucl. Acids Res. **37:**D211-215.

15. **Jenney, F. E., and M. W. W. Adams.** 2001. Rubredoxin from *Pyrococcus furiosus*. Methods Enzymol. **334:**45-55.

16. **Karthikeyan, S., U. M. Joshi, and R. Balasubramanian.** 2006. Microwave assisted sample preparation for determining water-soluble fraction of trace elements in urban airborne particulate matter: evaluation of bioavailability. Anal. Chim. Acta **576:**23-30.

17. **Kasampalidis, I. N., I. Pitas, and K. Lyroudia.** 2007. Conservation of metal-coordinating residues. Proteins: Struct. Funct. Bioinf. **68:**123-130.

18. **Kosnett, M. J.** 2007. Heavy metal intoxication and chelators, p. 945-957. *In* B. G. Katzung (ed.), Basic and clinical pharmacology (10th ed.). McGraw-Hill, New York

19. **Lobinski, R., D. Schaumlöffel, and J. Szpunar.** 2006. Mass spectrometry in bioinorganic analytical chemistry. Mass Spec. Rev. **25:**255-289.

20. **Luttringer, F., E. Mulliez, B. Dublet, D. Lemaire, and M. Fontecave.** 2009. The Zn center of the anaerobic ribonucleotide reductase from *E. coli*. J. Biol. Inorg. Chem. **14:**923-933.

21. **Menon, A. L., F. L. I. Poole, A. Cvetkovic, S. A. Trauger, E. Kalisiak, J. W. Scott, S. Shanmukh, J. Praissman, F. E. J. Jenney, W. R. Wilkoff, J. V. Apon, G. Siuzdak, and M. W. W. Adams.** 2009. Novel protein complexes identified in the hyperthermophilic archaeon *Pyrococcus furiosus* by non-denaturing fractionation of the native proteome. Mol. Cell. Proteomics **8:**735-751.

22. **Messerschmidt, A.** 2005. Handbook of Metalloproteins, vol. 1-3. John Wiley, NY.

23. **Perry, J. J., D. S. Shin, E. D. Getzoff, and J. A. Tainer.** 2010. The structural biochemistry of the superoxide dismutases. Biochim Biophys Acta **1804:**245-62.

24. **Poole, F. L., II, B. A. Gerwe, R. C. Hopkins, G. J. Schut, M. V. Weinberg, F. E. Jenney, Jr., and M. W. W. Adams.** 2005. Defining genes in the genome of the hyperthermophilic archaeon *Pyrococcus furiosus*: implications for all microbial genomes. Journal of Bacteriology **187:**7325-7332.

25. **Quevillon, E., V. Silventoinen, S. Pillai, N. Harte, N. Mulder, R. Apweiler, and R. Lopez.** 2005. InterProScan: protein domains identifier. Nucl. Acids Res. **33:**W116-120.

26. **Ragsdale, S. W.** 2009. Nickel-based Enzyme Systems. J Biol Chem **284:**18571-5.

27. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: A Laboratory Manual, 2 ed, vol. 3. Cold Spring Harbor Laboratory Press.

28. **Sanz-Medel, A., M. Montes-Bayón, M. del Rosario Fernández de la Campa, J. Encinar, and J. Bettmer.** 2008. Elemental mass spectrometry for quantitative proteomics. Analyt. Bioanalyt. Chem. **390:**3-16.

29. **Schut, G. J., S. L. Bridger, and M. W. W. Adams.** 2007. Insights into the Metabolism of Elemental Sulfur by the Hyperthermophilic Archaeon *Pyrococcus furiosus*: Characterization of a Coenzyme A- Dependent NAD(P)H Sulfur Oxidoreductase. Journal of Bacteriology **189:**4431-4441.

30. **Schwarz, G., R. R. Mendel, and M. W. Ribbe.** 2009. Molybdenum cofactors, enzymes and pathways. Nature **460:**839-47.

31. **Shi, W., C. Zhan, A. Ignatov, B. A. Manjasetty, N. Marinkovic, M. Sullivan, R. Huang, and M. R. Chance.** 2005. Metalloproteomics: high-throughput structural and functional annotation of proteins in structural genomics. Structure **13:**1473-1486.

32. **Shu, N., T. Zhou, and S. Hovmoller.** 2008. Prediction of zinc-binding sites in proteins from sequence. Bioinformatics **24:**775-782.

33. **Splan, K. E., K. Musier-Forsyth, M. T. Boniecki, and S. A. Martinis.** 2008. In vitro assays for the determination of aminoacyl-tRNA synthetase editing activity. Methods **44:**119-128.

34. **Waldron, K. J., J. C. Rutherford, D. Ford, and N. J. Robinson.** 2009. Metalloproteins and metal sensing. Nature **460:**823-830.

35. **Weinberg, M. V., G. J. Schut, S. Brehm, S. Datta, and M. W. W. Adams.** 2005. Cold shock of a hyperthermophilic archaeon: *Pyrococcus furiosus* exhibits multiple responses to a suboptimal growth temperature with a key role for membrane-bound glycoproteins. Journal of Bacteriology **187:**336-348.

36. **Zhang, Y., and V. N. Gladyshev.** 2010. General trends in trace element utilization revealed by comparative genomic analyses of Co, Cu, Mo, Ni, and Se. Journal of Biological Chemistry **285:**3393-405.

37. **Zillig, W., K. O. Stetter, S. Wunderl, W. Schulz, H. Preiss, and I. Scholz.** 1980. The *Sulfolobus*-"*Caldariella*" group: taxonomy on the basis of the structure of DNA-dependent RNA polymerases Arch. Microbiol. **125:**259-269.

Table C.1. Metalloproteins purified from *P. furiosus* by metal-based chromatography

Metal peak purified	Protein purified*	Metals present†	Proposed metal coordination	IPM-predicted metal	Annotated function‡
Mo	PF1972 (18978344)	Mo, Fe	$Mo(Cys)_4$	Fe	Radical SAM activase
Mo	PF1587 (18977959)	Mo	$Mo(Cys)_4$	None	Conserved hypothetical
Ni	PF0086 (18976458)	Ni	$Ni(His)_3Cys$	Zn	Alanyl-tRNA editing hydrolase
Ni	PF0056 (18976428)	Ni, Zn	$Ni(His)_3Glu$	None	Cupin/putative sugar-binding protein
U	PF0742 (18977114)	Fe (U, Pb)	–	Fe	Ferritin
U	PF0215 (18976587)	(U)	–	None	Enolase
Pb	PF1343 (18977715)	Zn (Pb)	–	Zn	Proline dipeptidase
Pb	PF0257 (18976629)	Fe (Pb)	–	None	Inorganic pyrophosphatase

*The NCBI GI number is given in parenthesis.
†Metals present in low amounts (<0.01 atoms mol^{-1}) are given in parenthesis.
‡On the basis of the annotation in the InterPro database.

Figure C.1 Metal assimilation by *P. furiosus* and unassigned metal peaks. a–c, Relative amounts (percentage on a molar basis) of the ten metals present as peaks in the C1 chromatography fractions (**a**) and those ten metals in the growth medium (**b**) (21), and number of metal peaks in the C2 chromatography fractions that can be assigned (solid bars) or cannot be assigned (shaded bars) to a protein with an InterPro-Metal (IPM) hit for that metal (**c**). The values for uranium are from the C1 column. The order of metals in the bar graphs reflects their abundance in the C1 fractions (the W content is 34% in the medium) and data for Fe and Zn are omitted for clarity. Metals that *P. furiosus* was not known to use are underlined (see Supplementary Tables 3 and 5).

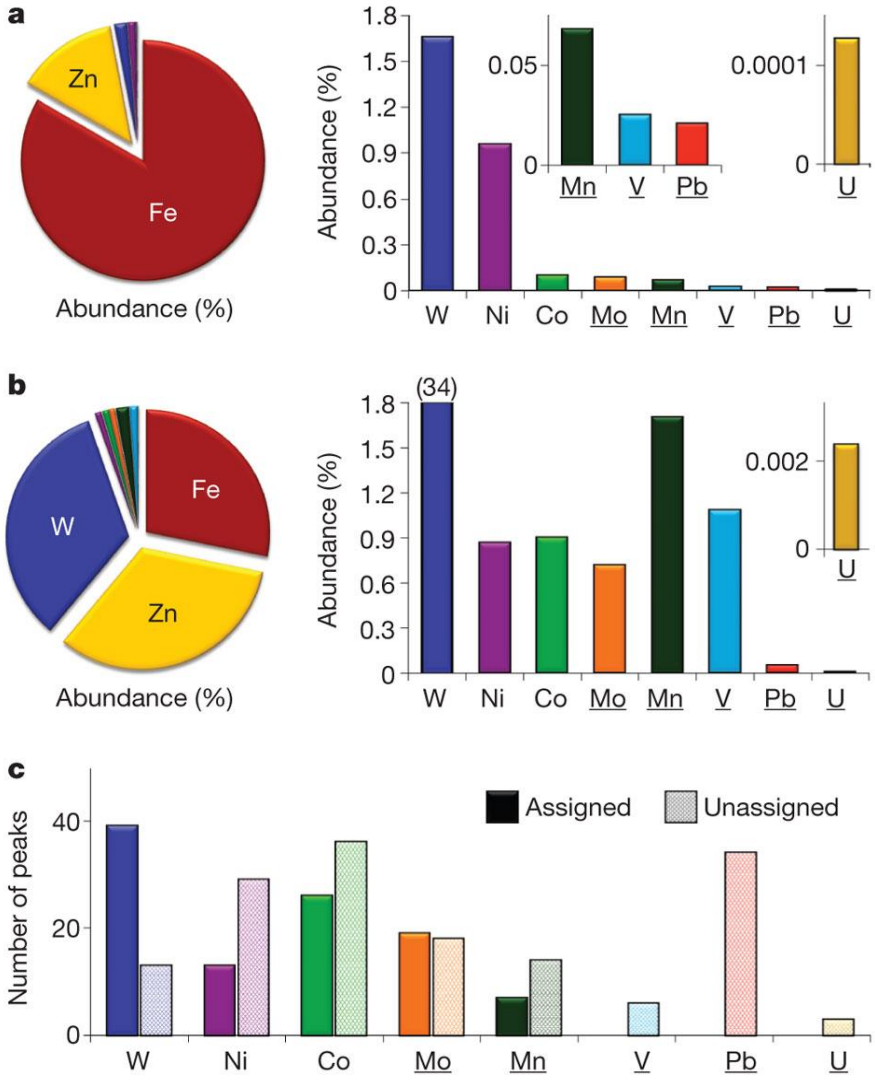
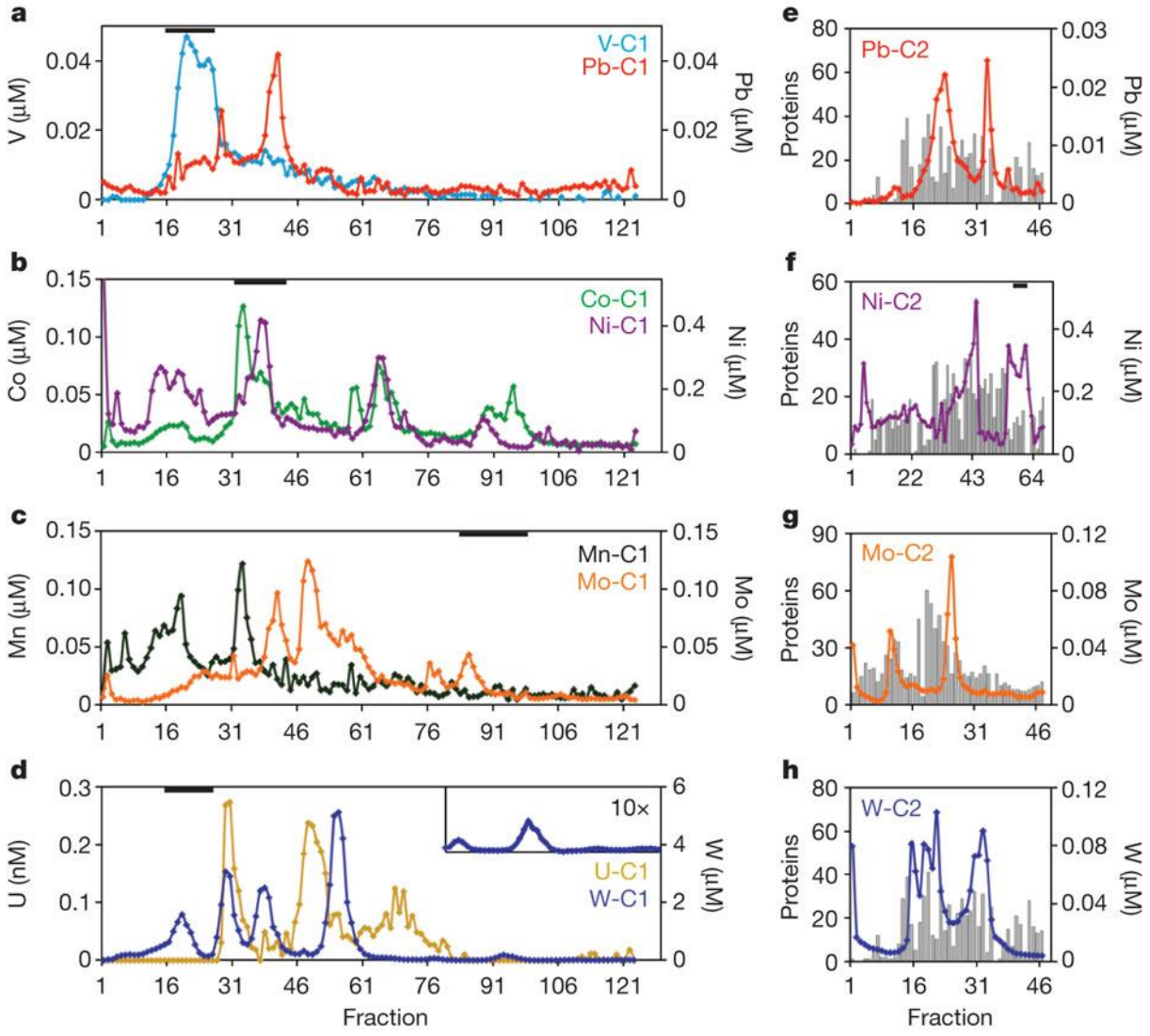

a
Zn
Fe
Abundance (%)
Abundance (%)
1.8
1.5
1.2
0.9
0.6
0.3
0
0.05
0
0.0001
0
W Ni Co Mo Mn V Pb U
Mn V Pb
U
b
Fe
W
Zn
Abundance (%)
(34)
0.002
c
Assigned
Unassigned
Number of peaks
40
20
0
W Ni Co Mo Mn V Pb U

Figure C.2 Metal concentration profiles after chromatographic fractionation *of P. furiosus* cytoplasmic extract. a–d, The C1 columns are vanadium (V) and lead (Pb) (**a**); nickel (Ni) and cobalt (Co) (**b**); molybdenum (Mo) and manganese (Mn) (**c**); tungsten (W) and uranium (U) (**d**). The bold lines above the fractions in the C1 columns indicate which were applied to a subsequent (C2) column. **e–h**, The metal concentrations and the number of proteins in the C2 columns are shown for Pb (**e**), Ni (**f**), Mo (**g**) and W (**h**). The bold line above the fractions in the Ni C2 column indicates which were applied to a subsequent (C3) column (see text and Supplementary Tables 10 and 11).


a
V-C1
Pb-C1
V (µM)
Pb (µM)
b
Co-C1
Ni-C1
Co (µM)
Ni (µM)
c
Mn-C1
Mo-C1
Mn (µM)
Mo (µM)
d
10×
U-C1
W-C1
U (nM)
W (µM)
Fraction
e
Pb-C2
Proteins
Pb (µM)
f
Ni-C2
Proteins
Ni (µM)
g
Mo-C2
Proteins
Mo (µM)
h
W-C2
Proteins
W (µM)
Fraction