

Microbial metalloproteomes are largely uncharacterized

Aleksandar Cvetkovic^{1*}, Angeli Lal Menon^{1*}, Michael P. Thorgersen^{1*}, Joseph W. Scott¹, Farris L. Poole II¹, Francis E. Jenney Jr^{1†}, W. Andrew Lancaster¹, Jeremy L. Praissman¹, Saratchandra Shanmukh¹, Brian J. Vaccaro¹, Sunia A. Trauger², Ewa Kalisiak², Junefredo V. Apon², Gary Siuzdak², Steven M. Yannone³, John A. Tainer³ & Michael W. W. Adams¹

Metal ion cofactors afford proteins virtually unlimited catalytic potential, enable electron transfer reactions and have a great impact on protein stability^{1,2}. 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^{2–4}. 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 (*Pyrococcus furiosus*). 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.

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^{1,2,5}. For example, unexpected zinc and iron–sulphur sites gave fundamental insights into DNA repair proteins relevant to human cancers⁶. Unfortunately, the small fraction of biochemically characterized proteins and limitations of metalloprotein bioinformatics^{1,4,5,7–9} 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^{8,10–13}. Yet, native biomass is likely essential as metalloproteins from recombinant sources may have incorrect or no metal at all^{14,15}.

Herein we present combined technologies that reveal assimilated metals and metalloproteins from biomass of the prototypical microbe *Pyrococcus furiosus*¹⁶ (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-MS¹¹ were detected in the cytoplasmic extract, including lead (Pb), titanium (Ti) and uranium (U) (Supplementary Tables 2 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 and Table 3). 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 environment¹⁷.

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)¹⁸. 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/MS¹⁸ identified 770 proteins or ~60% of the cytoplasmic proteins¹⁹. Given the difficulties of predicting metalloproteins^{1,4,5,7,9}, we searched the Integrated Resource of Protein Domains and Functional Sites (InterPro)²⁰ 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

¹Department of Biochemistry and Molecular Biology, University of Georgia, Athens, Georgia 30602, USA. ²Scripps Center for Mass Spectrometry and the Departments of Molecular Biology and Chemistry, The Scripps Research Institute, La Jolla, California 92037, USA. ³Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. †Present address: Philadelphia College of Osteopathic Medicine, Suwanee, Georgia 30024, USA (F.E.J.).

*These authors contributed equally to this work.

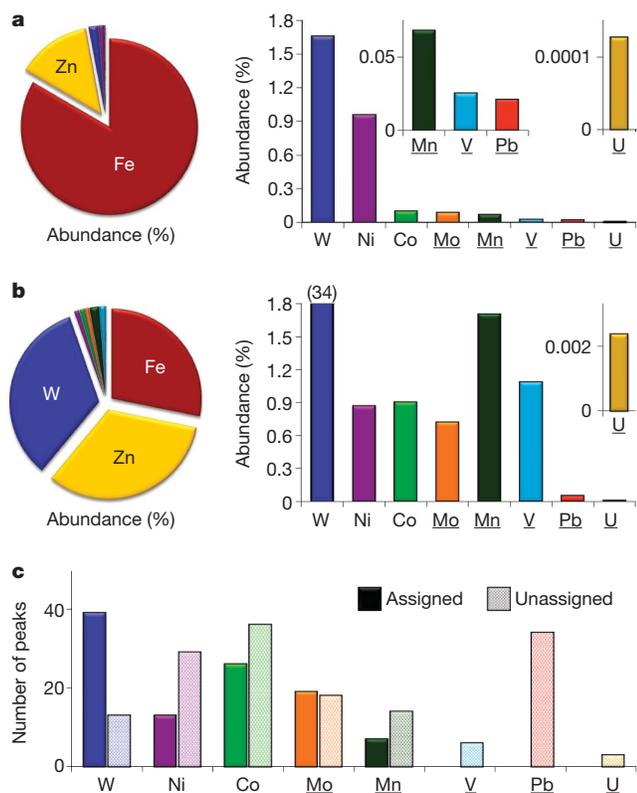


Figure 1 | Metal assimilation by *P. furiosus* and unassigned metal peaks. **a–c**, Relative amounts (percentage on a molar basis) of the ten metals present as peaks in the C1 chromatography fractions (a) and those ten metals in the growth medium (b)¹⁸, and number of metal peaks in the C2 chromatography 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).

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 superfamilies²¹; 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)²². 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 conditions²³. 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 cofactor²⁴. 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 biomass¹⁸ 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 Zn²⁵ 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 ten²⁶. Such Ni-enzyme discoveries can enhance understanding of catalysis and biology as seen for Ni versus other metal-containing superoxide dismutases²⁷.

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 (20.9 kDa; 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 Mg²⁺-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

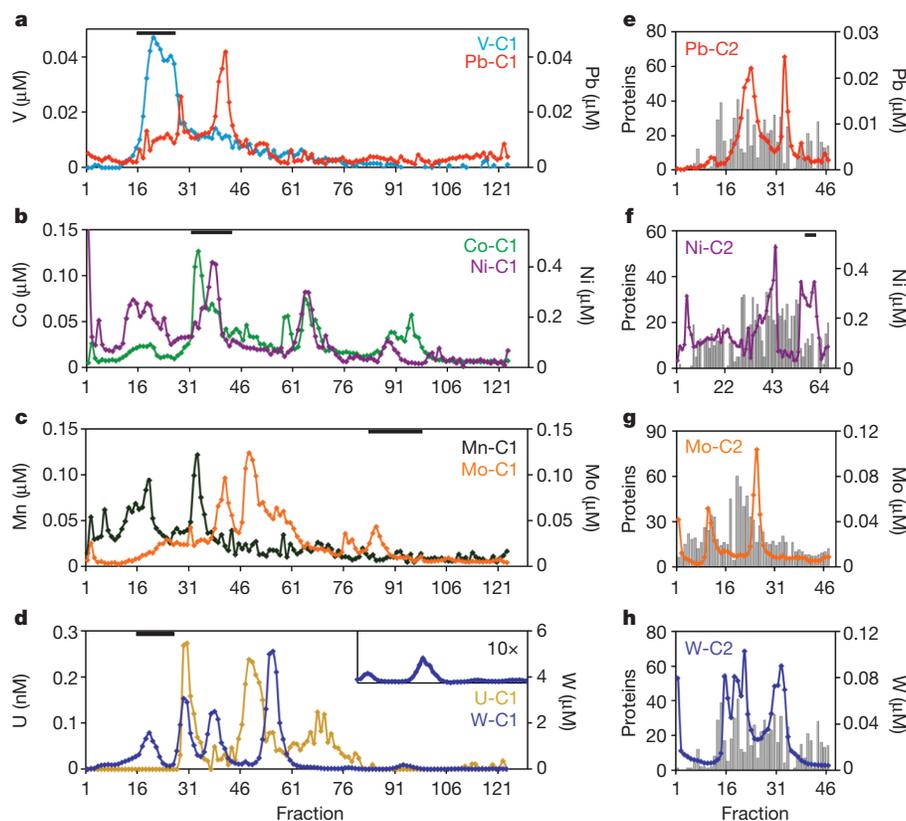


Figure 2 | Metal concentration profiles after chromatographic fractionation of *P. furiosus* cytoplasmic extract. **a–d**, The C1 columns are vanadium (V) and lead (Pb) (**a**); nickel (Ni) and cobalt (Co) (**b**); molybdenum (Mo) and manganese (Mn) (**c**); tungsten (W) and uranium (U) (**d**). The 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).

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*²⁸. 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. Notably, 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^{29,30}. The power and flexibility of this metal-based approach makes it a valuable

Table 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) ₃ Cys	Zn	Alanyl-tRNA editing hydrolase
Ni	PF0056 (18976428)	Ni, Zn	Ni(His) ₃ Glu	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.

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 phase¹⁸. 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 elsewhere¹⁸. 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 concentration¹⁸. 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 (ref. 28). Cytoplasmic extracts of each organism were subjected to anion exchange chromatography and metal analysis using the procedures devised for *P. furiosus*¹⁸. See Methods for details.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 8 April; accepted 7 May 2010.

Published online 18 July 2010.

- Gray, H. B., Stiefel, E. I., Valentine, J. S. & Bertini, I. *Biological Inorganic Chemistry: Structure and Reactivity* (Univ. Science Books, 2006).
- Messerschmidt, A., Huber, R., Wieghart, K. & Poulos, T. *Handbook of Metalloproteins*, Vol. 1–3. (Wiley, 2005).
- Shu, N., Zhou, T. & Hovmoller, S. Prediction of zinc-binding sites in proteins from sequence. *Bioinformatics* **24**, 775–782 (2008).
- Kasampalidis, I. N., Pitas, I. & Lyroudia, K. Conservation of metal-coordinating residues. *Proteins: Struct. Funct. Bioinf.* **68**, 123–130 (2007).
- Castagnetto, J. M. et al. MDB: the metalloprotein database and browser at the Scripps Research Institute. *Nucleic Acids Res.* **30**, 379–382 (2002).
- Fan, L. et al. XPD helicase structures and activities: insights into the cancer and aging phenotypes from XPD mutations. *Cell* **133**, 789–800 (2008).
- Andreini, C., Bertini, I., Cavallaro, G., Holliday, G. L. & Thornton, J. M. Metal-MACIE: a database of metals involved in biological catalysis. *Bioinformatics* **25**, 2088–2089 (2009).
- Waldron, K. J., Rutherford, J. C., Ford, D. & Robinson, N. J. Metalloproteins and metal sensing. *Nature* **460**, 823–830 (2009).
- Zhang, Y. & Gladyshev, V. N. General trends in trace element utilization revealed by comparative genomic analyses of Co, Cu, Mo, Ni, and Se. *J. Biol. Chem.* **285**, 3393–3405 (2010).
- Lobinski, R., Schaumlöffel, D. & Szpunar, J. Mass spectrometry in bioinorganic analytical chemistry. *Mass Spec. Rev.* **25**, 255–289 (2006).
- Sanz-Medel, A., Montes-Bayón, M., del Rosario Fernández de la Campa, M., Encinar, J. R. & Bettmer, J. Elemental mass spectrometry for quantitative proteomics. *Analyt. Bioanalyt. Chem.* **390**, 3–16 (2008).
- Shi, W. et al. Metalloproteomics: high-throughput structural and functional annotation of proteins in structural genomics. *Structure* **13**, 1473–1486 (2005).
- Atanassova, A., Högbom, M. & Zamble, D. B. in *Methods in molecular biology* Vol. 436 (eds B. Kobe, M. Guss & T. Huber) 319–330 (Humana Press, 2008).

- Jenney, F. E. & Adams, M. W. W. Rubredoxin from *Pyrococcus furiosus*. *Methods Enzymol.* **334**, 45–55 (2001).
- Chai, S. C., Wang, W. L. & Ye, Q. Z. Fe(II) is the native cofactor for *Escherichia coli* methionine aminopeptidase. *J. Biol. Chem.* **283**, 26879–26885 (2008).
- Fiala, G. & Stetter, K. O. *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaeobacteria growing optimally at 100 °C. *Arch. Microbiol.* **145**, 56–61 (1986).
- Aiuppa, A., Dongarra, G., Capasso, G. & Allard, P. Trace elements in the thermal groundwaters of Vulcano island (Sicily). *J. Volc. Geoth. Res.* **98**, 189–207 (2000).
- Menon, A. L. et al. Novel protein complexes identified in the hyperthermophilic archaeon *Pyrococcus furiosus* by non-denaturing fractionation of the native proteome. *Mol. Cell. Proteomics* **8**, 735–751 (2009).
- Poole, F. L. II et al. Defining genes in the genome of the hyperthermophilic archaeon *Pyrococcus furiosus*: implications for all microbial genomes. *J. Bacteriol.* **187**, 7325–7332 (2005).
- Hunter, S. et al. InterPro: the integrative protein signature database. *Nucleic Acids Res.* **37**, D211–D215 (2009).
- Agarwal, G., Rajavel, M., Gopal, B. & Srinivasan, N. Structure-based phylogeny as a diagnostic for functional characterization of proteins with a cupin fold. *PLoS ONE* **4**, e5736 (2009).
- Luttringer, F., Mulliez, E., Dublet, B., Lemaire, D. & Fontecave, M. The Zn center of the anaerobic ribonucleotide reductase from *E. coli*. *J. Biol. Inorg. Chem.* **14**, 923–933 (2009).
- Weinberg, M. V., Schut, G. J., Brehm, S., Datta, S. & Adams, M. W. W. 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**, 336–348 (2005).
- Schwarz, G., Mendel, R. R. & Ribbe, M. W. Molybdenum cofactors, enzymes and pathways. *Nature* **460**, 839–847 (2009).
- Splan, K. E., Musier-Forsyth, K., Boniecki, M. T. & Martinis, S. A. *In vitro* assays for the determination of aminoacyl-tRNA synthetase editing activity. *Methods* **44**, 119–128 (2008).
- Ragsdale, S. W. Nickel-based enzyme systems. *J. Biol. Chem.* **284**, 18571–18575 (2009).
- Perry, J. J., Shin, D. S., Getzoff, E. D. & Tainer, J. A. The structural biochemistry of the superoxide dismutases. *Biochim. Biophys. Acta* **1804**, 245–262 (2010).
- Zillig, W. et al. The *Sulfolobus*-“*Caldariella*” group: taxonomy on the basis of the structure of DNA-dependent RNA polymerases. *Arch. Microbiol.* **125**, 259–269 (1980).
- Kosnett, M. J. in *Basic and clinical pharmacology* 10th ed. (ed. B. G. Katzung) 945–957 (McGraw-Hill, 2007).
- Bressler, J. P. et al. Metal transporters in intestine and brain: their involvement in metal-associated neurotoxicities. *Hum. Exp. Toxicol.* **26**, 221–229 (2007).

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.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.W.W.A. (adams@bmb.uga.edu).

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 elsewhere¹⁸. 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(NO₃)₂, UO₂(C₂H₃O₂)₂·2H₂O, RhCl₃·3H₂O and RuCl₃·3H₂O, each at 50 nM) and chromium (CrCl₃·6H₂O, 200 nM). CMMex medium is complete maltose medium^{31,32} containing elemental sulphur (31 mM) to which Pb(NO₃)₂ and UO₂(C₂H₃O₂)₂·2H₂O were added at 500 nM each. Cells were processed for the C1 fractionation step as described for cells grown in the RM medium¹⁸. 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 100,000g 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 previously¹⁸. 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 MALDI time-of-flight mass spectrometer in reflectron mode using positive ion detection.

The mass list was generated by the SNAP peak detection algorithm using a signal-to-noise threshold of four following baseline correction of the spectra. Proteins were identified by searching the mass list against the National Center for Biotechnology Information (NCBI) annotation of the *P. furiosus* genome (NC_003413) using Mascot's Peptide Mass Fingerprint tool (version 2.1, Matrix Science). The searches were conducted using a peptide mass tolerance of 1.0, variable modifications of carbamidomethylation (C) and oxidation (M), and a maximum of one missed cleavage. Proteins with a *P* < 0.05 (corresponding to a Mascot protein score greater than 46) were considered significant.

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 ⁵⁶Fe, 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 HNO₃ (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 2,800g 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 analysis³³, the most commonly used acid, HNO₃, has the benefits of wide elemental solubility combined with low levels of interference and signal instability³⁴. Native purified *P. furiosus* rubredoxin (PF1282), a highly stable iron-containing metalloprotein³⁵, was used to develop the pretreatment procedure for ICP-MS analysis. The effect of HNO₃ 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) HNO₃ 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) HNO₃ 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. furiosus* genome using the 2007 InterProScan tool³⁶ and the InterPro (IPR) database²⁰. 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/MS¹⁸ 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, NiCl₂, CoCl₂ or ZnCl₂ (each 200 µM) to cells growing in NZCYM rich medium, which contains casein hydrolysate, casamino acids and yeast extract³⁷. 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'-AAGCTCGAGCGGCCGCTAATCTTCCAGCCATATCTCCAATC-3' containing an XhoI restriction site (restriction sites are underlined). The PCR product was digested with the restriction enzymes, NdeI and XhoI, 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 A₆₀₀ 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, NiCl₂, CoCl₂ or ZnCl₂ 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 48,000g 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 paste³⁷. 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 47,000g 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 paste²⁸. 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).

31. Adams, M. W. W. *et al.* Key role for sulfur in peptide metabolism and in regulation of three hydrogenases in the hyperthermophilic archaeon *Pyrococcus furiosus*. *J. Bacteriol.* **183**, 716–724 (2001).
32. Schut, G. J., Bridger, S. L. & Adams, M. W. W. Insights into the metabolism of elemental sulfur by the hyperthermophilic archaeon *Pyrococcus furiosus*: characterization of a coenzyme A-dependent NAD(P)H sulfur oxidoreductase. *J. Bacteriol.* **189**, 4431–4441 (2007).
33. Cai, Y., Georgiadis, M. & Fourqurean, J. W. Determination of arsenic in seagrass using inductively coupled plasma mass spectrometry. *Spectrochim. Acta B* **55**, 1411–1422 (2000).
34. Karthikeyan, S., Joshi, U. M. & Balasubramanian, R. 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 (2006).
35. Blake, P. R. *et al.* 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 (1991).
36. Quevillon, E. *et al.* InterProScan: protein domains identifier. *Nucleic Acids Res.* **33**, W116–W120 (2005).
37. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular cloning: A Laboratory Manual* 2nd ed. Vol. 3 (Cold Spring Harbor Laboratory Press, 1989).