**FINAL PROGRESS REPORT**

**MAGGIE Component 1: Identification and Purification of Native and Recombinant Multiprotein Complexes and Modified Proteins from *Pyrococcus furiosus***

**Director:** Michael W. W. Adams, University of Georgia

**DOE Award:** DE-FG02-07ER64326

**Award Dates:** 11/01/2007 – 10/31/2011

## Background and Significance

Virtualy all cellular processes are carried out by dynamic molecular assemblies or multiprotein complexes (PCs), the composition of which is largely unknown. The results from recent structural genomics efforts have clearly demonstrated that less than 25% of the open-reading-frames (ORFs) in a given prokaryotic genome will yield stable, soluble proteins when expressed using a one-ORF-at-a-time approach. We would argue that the remaining 75% of the ORFs encode proteins that are part of multiprotein complexes (PCs) or are modified post-translationally, e.g. with metals, such that the recombinant form from a single ORF is unstable and cannot be purified. The problem is that PCs and metalloproteins (MPs) cannot be accurately predicted on a genome-wide scale. As more genome sequence information becomes available, the situation is becoming more critical. For example, when the original MAGGIE proposal was written (Jan., 2005) there were 244 complete genome sequences, now (January, 2014) there are >12,000 genome projects completed.

How can PCs and MPs be determined on a genome-wide scale? They cannot be predicted solely from bioinformatics analyses, nor are there well-defined experimental techniques currently available to unequivocally identify them. While there are several methods available to study protein-protein interactions on a genome-wide scale, a major drawback is that they require genetic manipulation of the target organism, an approach limited to only a few very well-studied systems. Similarly, non-genetic techniques to identify protein-protein interactions are limited in their dynamic range and typically identify only high abundance proteins. The only solution to this dilemma is to experimentally determine PCs and MPs in biomass of a model organism and to develop analytical tools that can then be applied to the biomass of any other organism. In other words, organisms themselves must be analyzed to identify their PCs and MPs: "native proteomes" must be determined. This information can then be utilized to design multiple ORF expression systems to produce recombinant forms of PCs and MPs. Moreover, the information and utility of this approach can be enhanced by using a hyperthermophile, one that grows optimally at 100°C, as a model organism. By analyzing the native proteome at close to 100 °C below the optimum growth temperature, we will trap reversible and dynamic complexes, thereby enabling their identification, purification, and subsequent characterization. The model organism for the current study is *Pyrococcus furiosus,* a hyperthermophilic archaeon that grows optimally at 100°C. It is grown up to 600-liter scale and kg quantities of biomass are available.

The Component 1 goals in the original proposal were (1) to identify native PCs and MPs using native *P. furiosus* biomass (with MS/MS analyses to identify proteins by component 4), (2) to provide samples of abundant native PCs and MPs for structural characterization (using SAXS by component 5), (3) to obtain recombinant versions of the less abundant PCs and MPs using multiple gene expression systems with homologous and heterologous genetic approaches, and (4) to design and evaluate generic

protocols for PC and MP production in other prokaryotes of DOE interest.

## Progress and Accomplishments

**Cytoplasmic Multiprotein Complexes (PCs).** We have directly determined the identity of PCs from native microbial biomass of *P. furiosus*. A new high throughput MS/MS approach was developed with component 4 to accommodate the large number of fractions that were analyzed. The results were published in *Molecular Cellular Proteomics*. A total of 1,003 distinct proteins were identified by nano LC-ESI-MS/MS, which represent approximately 80% of the cytoplasmic proteins. Based on the co-fractionation of proteins that are encoded by adjacent genes on the chromosome, 106 potential heteromeric PCs containing 243 proteins were identified, only 20 of which were known or expected. The list of 106 PCs was made available to the GTL and scientific community (http://masspec.scripps.edu/maggie/).

**Structural characterization of *P. furiosus* PCs by Component 5.** PCs purified from native biomass of *P. furiosus* and recombinant *P. furiosus* proteins produced in *E. coli* have been supplied to Component 5 for structural analyses by dynamic light scattering and small-angle X-ray scattering (SAXS). This resulted in a paper published in *Nature Methods*.

**Comparative Transcriptomics, Metabolomics and Metallomics.** In collaboration with components 1, 2, 3 and 4, experiments was designed to study the effect of growth phase on the transcriptome, metabolome and metallome of *P. furiosus*, *S. solfataricus*, *Halobacterium* and *M. maripaludis*.

**Metalloproteins (MPs).** Metal ion co-factors afford proteins virtually unlimited catalytic potential, enable electron transfer reactions and are major determinants of protein stability. Consequently, metalloproteins play key roles in most biological processes, including respiration (iron and copper), photosynthesis (manganese) and drug metabolism (iron). Yet, the metalloproteome of any single organism has yet to be defined. It is not currently possible computationally to predict from a genome sequence the types of metal that an organism takes up from its environment or the numbers and types of metalloprotein it contains because metal coordination sites are diverse and not easily recognized. We have developed a generally-applicable, metal-based approach is described to determine all metals assimilated by an organism and to identify metalloproteins on a genome-wide scale. This approach shifts the focus from traditional protein purification to metal identification and purification. We employ liquid chromatography, high-throughput tandem mass spectrometry (HT-MS/MS) and inductively coupled plasma mass spectrometry (ICP-MS) to separate and identify metalloproteins. After fractionation of a cytoplasmic extract from a model microorganism (*Pyrococcus furiosus*), 343 metal peaks were detected, 158 of which represented no known or predicted metalloprotein. These unassigned peaks included metals known to be utilized by the organism (cobalt, iron, nickel, tungsten and zinc) as well as metals it was not known to assimilate (lead, manganese, molybdenum, uranium and vanadium). Eight unassigned metal peaks were purified yielding four completely novel Ni- and Mo-containing proteins, and four proteins containing low amounts of lead and uranium, metals that appear to be inadvertently inserted and potentially the basis for metal toxicity. Metalloproteomes involving both biologically-conventional and unanticipated metals are therefore much more extensive and diverse than previously recognized, with implications for a more complete understanding of both cell biology and metal toxicity. The results of this work were submitted. This resulted in a paper published in *Nature.*

**Metal Content of and Novel Metalloproteins in other Microorganisms.** In collaboration with component 2, 600-liter fermentations of *S. solfataricus* were grown (at pH 3, 75°C) at the University of Georgia yielding 1.5 kg of biomass. The elemental composition of native biomass of two bacteria, *Escherichia coli* (facultative aerobe, mesophile, bacterium) and the bioenergy-relevant *Anaerocellum thermophilum* (anaerobe, cellulolytic thermophile, bacteria) were also analzyed. A total of 28 (of 52 elements analyzed) were detected in one of more of the four organisms, with 6 elements (including Pb) found in all four, and 4 elements (including U) in three.

**Peer-Reviewed Papers (resulting from this DOE-funded research)**


Trauger, S. A., Kalisak, E., Kalisiak, J., Morita, H., Weinberg, M.V., Menon, A. L., Poole, F. L., Adams, M. W. W. and Siuzdak, G. (2008) Correlating the transcriptome, proteome, and metabolome in the environmental adaptation of a hyperthermophile. *J. Proteome. Res*. **7,** 1027-1035

Menon, A. L., Poole, F. L., Cvetkovic, A., Trauger, S. A., Kalisiak, E., Scott, J. W., Shanmukh, S., Praissman, J., Jenney, F. E., Wikoff, W. R, Apon, J. V., Siuzdak, G. and Adams, M. W. W. (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

Hura, G. L., Menon, A. L., Hammel, M., Rambo, R., Poole, F. L., Tsutakawa, S., Jenney, F. E., Classen, S., Frankel, L., Hopkins, R. C., Yang, S.-J., Scott, J. W., Dillard, B. D., Adams, M. W. W. and Tainer, J. A. (2009) "Rapid and robust proteomics-scale structural analyses determined by x-ray solution scattering (SAXS)" *Nature Methods* **6**, 606 – 612

Kalisiak, J., Trauger, S. A., Kalisiak, E., Morita, H., Fokin, V. V., Adams, M. W. W., Sharpless, K. B. and Siuzdak, G. (2009) "Identification of a new endogenous metabolite and the characterization of its protein interactions through an immobilization approach" *J. Am. Chem. Soc.* **131,** 378-386

Cvetkovic, A., Menon, A. L., Thorgersen, M., Scott, J. W., Poole, F. L., Jenney, F. E., Lancaster, W. A., Praismann, J. L., Shanmukh, S., Vaccaro, B., Trauger, S. A., Kalisiak, E., Apon, J. V., Siuzdak, G., Yannone, S. M., Tainer, J. A. and Adams, M. W. W. (2010) "Microbial metalloproteomes are largely uncharacterized" *Nature* **466,** 779-782

Lancaster, W. A., Praismann, J. L., Poole, F. L., Cvetkovic, A., Menon, A. L., Scott, J. W., Poole, F. L., Jenney, F. E., Thorgersen, M., Kalisiak, E., Apon, J. V., Trauger, S. A., Siuzdak, G., Tainer, J. A. and Adams, M. W. W. (2011) "A computational framework for proteome-wide pursuit and prediction of metalloproteins using ICP-MS and MS/MS data" *BMC Bioformatics* **12,** 64

Holden, J. F., Menon, A. L. and Adams, M. W. W. (2011) "Hyperthermophile-metal interactions in hydrothermal vent environments" In: *Microbial Metal and Metalloid Metabolism: Advances and Applications,* Stolz J. F. and Ormeland, R. S., eds., pp. 39-62, ASM Press, Washington, DC

Yoon, S. H., Reiss, J. R., Bare, J. C., Tenenbaum, D., Pan, M., Slagel, J., Lim, S., Hackett, M., Menon, A. L., Adams, M. W. W., Yannone, S. M., Leigh, J. A., and Baliga, N. S. (2011) "Parallel evolution of transcriptome structure during genome reorganization" *Genome Research* **21,** 1892-1904

Yannone, S. M., Hartung, S., Menon, A. L., Adams, M. W. and Tainer, J. A. (2011) "Metals in biology: defining metalloproteomes" *Curr. Opin. Biotechnol.* **23,** 89-95

Vaccaro, B. J., Menon, A. L., Lancaster, W. A. and Adams, M. W. W. (2012) "Metallomics using Inductively Coupled Plasma Mass Spectrometry" *Current Protocols in Chemical Biology* **4,** 249-274