Section 4

Molecular Interactions

32

Exploring the Genome and Proteome of *Desulfitobacterium hafniense* DCB2 for its Protein Complexes Involved in Metal Reduction and Dechlorination

James M. Tiedje[1]* (tiedjej@msu.edu), John Davis[2], Sang-Hoon Kim[1], David Dewitt[1], Christina Harzman[1], Robin Goodwin[1], Curtis Wilkerson[1], Joan Broderick[1,3], and Terence L. Marsh[1]

[1]Michigan State University, East Lansing, MI; [2]Columbus State University, Columbus, GA; and [3]Montana State University, Bozeman, MT

Desulfitobacterium hafniense is an anaerobic, low GC, Gram-positive spore-forming bacterium that shows considerable promise as a bioremediative competent population of sediments. Of particular relevance to its remediation capabilities are metal reduction, which changes the mobility and toxicity of metals, and chlororespiration, the ability to dechlorinate organic xenobiotics. Our work focuses on these two capabilities in *D. hafniense* DCB2 whose genome was recently sequenced by JGI and annotated by ORNL. We are determining the complete metal reducing repertoire of *D. hafniense* and the genes required for respiratory and non-respiratory metal reduction through DNA microarrays produced by ORNL and proteomics. In addition, genomic analysis identified seven putative reductive dehalogenases (RDases) and we are investigating these with genetic and biochemical tools. The status of these efforts is presented below.

Physiology

D. hafniense is capable of reducing iron as well as uranium, selenium, copper and cobalt. With the exception of selenium, the metals were reduced under conditions where the metal was the only available electron acceptor. Selenium was unable to serve in this capacity but was reduced when grown fermentatively. Grown in the presence of Se, the cellular morphology viewed with light microscopy and confirmed with SEM & TEM was elongated with small polyp-like spheres present on the outer surface and in the medium. Energy dispersive spectroscopy (EDS) localized the selenium to the polyps. Staining of the cells with a lipophilic fluorescent dye and osmium profiles from EDS scans suggest that the polyps are surrounded by a membrane. Similar metal localization studies are planned with uranium, copper, iron, and cobalt. Preliminary analysis of gene expression in fermentatively grown cells in the presence vs. the absence of selenium revealed two highly up-regulated genes coding for a radical SAM type protein (Figure 1) and a response regulator consisting of a CheY-like receiver domain and a HTH DNA-binding domain. Expression arrays are underway for identifying genes differentially expressed for the respiratory reduction of iron, uranium, copper and cobalt.

Proteomics

Protein identifications were carried out using 1- and 2-D gel separations of membrane and soluble fractions of *D. halfniense.* Bacteria labeled with ^{15}N are now being grown under fermentative conditions and will be compared to cells grown in ^{14}N in the presence of selenium, uranium, and iron. Sol-

* Presenting author

uble and membrane proteins isolated from mixtures of these *D. halfniense* cultures will be separated by SDS-PAGE and subjected to MS analysis on a ThermoFinnigan LTQ-FT mass spectrometer. The resulting spectra will be analyzed with X! Tandem to identify *D. halfniense* proteins. Comparison of the $^{15}N/^{14}N$ peak ratios are being use to quantify the differential expression of the *D. hafniense* proteins present in the fermentative, selenium-, uranium- and iron-grown cultures, and to correlate these changes with changes in the gene expression profile determine from microarray analysis.

Genetics

Of the seven putative RDase ORFs, two (*rdh1A* and *rdh3A)* were predicted to be nonfunctional due to a nonsense mutation and insertional disruption by a transposase gene, respectively. In each case, the non-functionality was removed through PCR-based procedures resulting in ostensibly full-length RDase genes with 527-aa and 503-aa protein products, respectively, instead of truncated products of 345-aa and 352-aa. Cloning and expression of the altered RDase genes along with neighboring *rdhBC* genes (e.g. *rdh1ABC* and *rdh3ABC*) in *Escherichia coli* resulted in lethality to the host in both cases. Introduction of the *rdh1AB* (without *C*) genes were also apparently lethal. However, *E. coli* clones were obtained for the unaltered *rdh1ABC* cluster and isolated *rdh1C* genes suggesting that expression of the altered *rdh1A* gene is lethal to the *E. coli* host. We are investigating the causes of lethality and the activities of the full-length gene products since this electron accepting pathway may have a central role in the desired and versatile reductive properties of this species.

Biochemistry

The seven RDase genes of *D. hafniense* DCB-2 are being cloned into vectors for expression in *E. coli*. CprA has been cloned into pET44a+ and the resulting plasmid transformed into *E. coli* Rosetta blue cells. Addition of IPTG to the culture at log phase results in significant overexpression of a protein with an apparent molecular mass of approximately 50 kDa. The overexpressed protein has been identified by mass spectral analysis as CprA. It is somewhat soluble under lysis conditions, and has been partially purified with gel filtration chromatography. Both the crude lysate and the partially purified protein exhibit a faint brown coloration, consistent with the presence of an iron-sulfur cluster cofactor.

Based on initial microarray studies, gene 3921 (Contig. 809) is one of only two genes found to be strongly up-regulated under selenate-reducing conditions. This gene encodes a hypothetical protein of 35.8 kDa and comparative sequence analysis suggests the protein likely belongs to the Radical SAM protein superfamily, a group of proteins that utilize iron-sulfur clusters and S-adenosylmethionine to initiate biological radical reactions (Figure 1). Functional studies are currently underway.

Figure 1. Alignment blocks in Radical-SAM proteins. (Ref. Sofia et al., Nucl. Acid. Res. 2001, 29, 1097-1106.)

I. Cluster block (CX_3CX_2C highly conserved)
consensus: V/I ___G CN__RC__YCY __.
gene 3921: IGALNSCPNGCKYCYAN
II. SAM-binding block (Gly-rich, less conserved)
consensus: V_FTGGEPLL .__
gene 3921: IAAKYGIPLQTC
III. Third block (least conserved in superfamily)
consensus: E__LEAIK _L__E_G
gene 3921: ESPLLIGRLKPSDN

33

Comparison of Conserved Protein Complexes Across Multiple Microbial Species to Evaluate High-Throughput Approaches for Mapping the Microbial Interactome

D. Pelletier[1]* (pelletierda@ornl.gov), G. Hurst[1], S. Kennel[1], L. Foote[1], P. Lankford[1], T. Lu[1], W. McDonald[1], C. McKeown[1], J. Morrell-Falvey[1], D. Schmoyer[1], E. Livesay[3], F. Collart[2], D. Auberry[3], K. Auberry[3], Y. Gorby[3], B. Hooker[3], E. Hill[3], C. Lin[3], P. Moore[3], R. Moore[3], R. Saripalli[3], K. Victry[3], V. Kery[3], S. Wiley[3], and **M. Buchanan**[1]

[1]Oak Ridge National Laboratory, Oak Ridge, TN; [2]Argonne National Laboratory, Chicago, IL; and [3]Pacific Northwest National Laboratory, Richland, WA

The ever-increasing number of available complete genome sequences for bacteria that have been less extensively studied than model systems such as *E. coli* has led to the need for additional methods for system wide functional genomic analysis. Methods (2-hydrid or affinity isolation) for the systematic identification of protein-protein interaction networks have been applied to a number of organisms (e.g., yeast, *C. elegans*, *E. coli*, etc.). It is clear that no single method can be used to fully describe all the interactions in an organism. These data when combined with other genome scale analyses, the transcriptome, the proteome and phenotypic observations, can lead to a better understanding of molecular interactions involved in a biological process.

We have developed a general, scaleable methodology for mapping protein-protein interaction networks, based on affinity isolation of protein complexes, which can be extended to bacterial systems beyond well-characterized model organisms. Specific open reading frames from the annotated genome are identified, cloned and expressed as fusion protein "baits" bearing affinity tags in the target organism (endogenous approach) or in *E. coli* followed by purification then exposed to the proteins from wild type organisms (exogenous approach). Following affinity isolation, protein interactors are identified using mass spectrometry. The metadata are managed using Nautilus LIMS. MS data analyzed by SEQUEST and then undergo computational and statistical analysis. We are using this approach, as part of the Center for Molecular and Cellular Systems, on two phylogenetically distinct bacteria of interest to DOE—the alpha-proteobacterium *Rhodopseudomonas palustris*, and the gamma-proteobacterium *Shewanella oneidensis*. Initial experiments have focused on a number of protein complexes conserved across multiple microbial species to validate this approach for mapping microbial interactomes.

As an example of these studies, the results of mapping the interaction subnetworks for homologs of open reading frames annotated to be part of DNA-directed RNA polymerase, F1F0-ATP synthase and the RNA degradosome will be presented. Results from *R. palustris* and *S. oneidensis* will be compared to results obtained for homologous subnetworks in other organisms that are available in public protein interaction databases (BIND, DIP, etc.). Our data demonstrate the feasibility of our high-throughput approaches for mapping microbial protein interaction networks.

An internal project database and website tracks and summarizes data from all experiments. Publicly accessible views of selected pages of this website show a summary of the status of targeted proteins for the endogenous approach (http://maple.lsd.ornl.gov/cgi-bin/gtl_demo/public_target_status.cgi) and for the exogenous approach (http://maple.lsd.ornl.gov/cgi-bin/gtl_demo/public_ex_target_status.cgi,), and example interactor identifications for a selected subnetwork (http://maple.lsd.ornl.gov/gtl_demo/index.html).

* Presenting author

34

Advanced Technologies for Identifying Protein-Protein Interactions

R. Hettich[1]* (hettichrl@ornl.gov), G. Hurst[1], W. McDonald[1], H. Connelly[1], D. Pelletier[1], C. Pan[1], N. Samatova[1], G. Kora[1], V. Kertesz[1], S. Gaucher[2], T. Iqbal[2], M. Hadi[2], M. Young[2], G. Orr[3], M. Romine[3], D. Panther[3], S. Reed[3], D. Hu[3], E. Livesay[3], B. Hooker[3], S. Wiley[3], S. Kennel[1], and **M. Buchanan**[1]

[1]Oak Ridge National Laboratory, Oak Ridge, TN; [2]Sandia National Laboratories, Livermore, CA; and [3]Pacific Northwest National Laboratory, Richland, WA

The major focus of the **Center for Molecular and Cellular Systems (CMCS**) is the establishment of a *high-throughput identification pipeline for measuring protein-protein interactions*. In the development and operation of this pipeline, several needs have been uncovered that, if addressed, would provide dramatic improvements in overall performance. In particular, improvements are needed in measurement sensitivity, enhanced accuracy, speed of identification, more comprehensive characterizations, and information management tools. The impetus for development of advanced technologies is to alleviate current bottlenecks of the pipeline as well as continue to pursue state-of-the-art technologies for enhanced measurement throughput and data quality. Recent research has focused on key developments in techniques for more comprehensively examining protein interactions, improving LC-MS throughput, applying automated detection of FRET using flow cytometry, and developing computational tools for data mining.

For more extensive characterization of the molecular details of protein interactions, MS methodologies have been developed to extract information from *intact proteins* by either direct top-down MS measurements or by chemical cross-linking of proteins with MS detection of their proteolytic peptides. As a demonstration of the top-down MS approach, tandem affinity purifications were conducted for the Gln family of proteins from *R. palustris*. These proteins are key to sensing internal cellular ammonium levels and transducing the signal. Under ammonium-rich conditions, the proteins GlnK1, GlnK2, and GlnB are unmodified; however, under ammonium-starvation conditions, these proteins are modified by uridylylation and activate the AmtB ammonium transporter and the glutamine synthetase enzyme that are important for nitrogen fixation. Top-down MS measurements provided not only details about the level of uridylyation, but also verified the presence of multiple isoforms of each protein. Chemical crosslinking, used in conjuction with enzymatic digestion and LC/MS/MS, has potential as an important tool for probing protein-protein interactions. Data interpretation is currently a challenge because the "rules" for crosslinked peptide dissociation have not been well studied. We are thus systematically investigating the dissociation pathways open to crosslinked peptides using a series of defined inter- and intra-molecular crosslinked species. Dissociation pathways unique to crosslinked peptides have been found, and are being incorporated into our MS2Assign software. This will allow for more fully automated analysis of crosslinked peptide sequence, a prerequisite for high throughput experiments.

Because of the complexity of protein-protein interactions in microbial systems, it is essential to continue development of methodologies for high throughput measurements. While LC-MS/MS shows tremendous potential for the identification and characterization of protein-protein interactions, the sequential chromatographic separations are either a current or soon to be rate-limiting step for throughput. In order to address this issue, we have undertaken the development of a protocol for fast LC-MS/MS based on a Q-ToF mass spectrometry platform. This setup includes several HPLC columns connected in parallel for multiplexing the runs and solvent washes. By conducting consecutive runs with 30-min. chromatographic separations, we have been able to examine almost

30 samples in a 24-hour period. The reproducibility and sensitivity are comparable to the longer, more conventional LC-MS/MS methodologies, in which about 12 samples can be analyzed in a 24-hour period. The power of fluorescence resonance energy transfer (FRET) for detecting molecular interactions between proteins can be applied to high throughput screening of protein-protein interactions. Using automated cloning and fluorescence detection approaches, FRET can be used to screen a large number of protein pairs, tagged with fluorescent proteins or fluorescent antibodies. We have combined Gateway compatible cloning with the application of flow cytometry for high throughput FRET analysis in *Shewanella* to leverage existing molecular reagents. Cyan and Yellow fluorescent proteins were successfully expressed in *Shewanella* and were used as donor and acceptor fluorophores in a FRET pair. In addition, we have used FRET analysis between fluorescent antibodies, directed against *Shewanella* periplasmic proteins. We demonstrate that detection of FRET by flowcytometry enables the rapid screening of a large number of cells.

One of the biggest hurdles in the measurements of protein interactions continues to be limitations in computational approaches for data mining. Recent effort has been given to the development of computational methods for integrating top-down and bottom-up MS analyses ("PTMsearchPlus") and deciphering isotopically labeled samples for protein quantification ("ProRata"). In order to automate the interpretation of the top-down MS data, the algorithm "PTMsearchPlus" has been created and tested with several data sets. This program integrates both top-down and bottom-up MS data to exploit the advantages of each approach. The ability to combine high resolution molecular mass data on the intact proteins with the more extensive sequence coverage of the proteolytic peptides from the bottom-up data provides a powerful platform for not only identifying proteins, but also characterizing the presence of post-translational modifications and isoforms. Protein quantification is very important for not only measuring the amount of protein present, but also the elucidation of the stoichiometry of proteins present in complexes. Stable isotope labeling is one of the most widely employed methods for quantification measurements; however, current computational tools to extract the information are quite remedial. We have developed a new tool, "ProRata," which addresses not only the extraction of isotopic information from the labeled samples, but also provides confidence interval scoring for each protein identification.

Research sponsored by the Genomics:GTL program, Office of Biological and Environmental Research, U.S. Department of Energy, under contract No. DE-AC05-00OR22725 with Oak Ridge National Laboratory, managed and operated by UT-Battelle, LLC.

* Presenting author

35

Complementary Assays for Validating Protein Interactions Identified by High-Throughput Screening Techniques

Mitchel J. Doktycz[1]* (doktyczmj@ornl.gov), Jennifer L. Morrell-Falvey[1], W. Hayes McDonald[1], Gurusahai Khalsa-Moyers[1], Dale A. Pelletier[1], Stephen J. Kennel[1], Vladimir Kery[2], Galya Orr[2], Dehong Hu[2], Margaret F. Romine[2], David J. Panther[2], Brian S. Hooker[2], H. Steven Wiley[2], and Michelle V. Buchanan[1]

[1]Oak Ridge National Laboratory, Oak Ridge, TN and [2]Pacific Northwest National Laboratory, Richland, WA

Directed experimental validation of protein-protein interactions is essential for confirming associations identified by the Center for Molecular and Cellular Systems' (CMCS) high throughput pipeline and/or predicted by computational techniques. The CMCS core pipeline employs the methods of affinity purification and mass spectroscopy to discover protein interactions in cells. To complement this capability and improve the confidence of protein interaction measurements, the CMCS is also utilizing a combination of live cell and *in vitro* techniques to validate the interaction between specific pairs of proteins. These assays use a common cloning platform that is shared with the core pipeline. The live cell-based approaches include a co-localization assay, fluorescence resonance energy transfer measurements, and yeast two hybrid based tests. Complementing these assays are surface plasmon resonance (SPR) measurements for characterizing protein-protein associations *in vitro*. Use of a suite of technologies will aid in comprehensively identifying protein interactions in a cell. Model complexes are being used to assess the ability of these techniques to characterize transient interactions, low-copy number proteins, or membrane-associated proteins. In addition to confirming physical associations between proteins, these validation assays can provide important functional data on the spatial, temporal, and biophysical context of the interactions. Results from these complementary assays will be shown for protein interactions elucidated from the CMCS core pipeline and/or predicted computationally. Using a combination of methods to assess protein interactions provides the biological community with the high quality data required for understanding the myriad processes that occur in a cell.

* Presenting author

36

Computational Approaches for Aggregating and Scoring Protein-Protein Interaction Data

William R. Cannon[1]* (william.cannon@pnl.gov), W. Hayes McDonald[2], Don S. Daly[1], Denise Schmoyer[2], Gregory B. Hurst[2], Manesh B. Shah[2], Brian S. Hooker[1], Vladimir Kery[1], Stephen J. Kennel[2], H. Steven Wiley[1], and **Michelle V. Buchanan**[2]

[1]Pacific Northwest National Laboratory, Richland, WA and [2]Oak Ridge National Laboratory, Oak Ridge, TN

The GTL CMCS informatics pipeline for *Shewanella oneidensis* MR-1 and *Rhodopseudomonas palustris* protein-protein interaction networks melds information from two sources: data collected internally, including both high-throughput pipeline (see companion abstract: "Comparison of conserved protein complexes across multiple microbial species to evaluate high-throughput approaches for mapping the microbial interactome") and more targeted validation (see companion abstract: "Complementary Assays for Validating Protein Interactions Identified by High-throughput Screening Techniques") assays, and external data sources that both identify interactors and add biological context. A key function of this integrated pipeline is to assess the quality of new data and the processing pipeline itself using: targeted experiments, QC standards, technical and biological replicates, statistical modeling, and statistical process control. Protein interaction networks are then inferred using two different methods. A Bayesian method gives a global analysis of the entirety of the data including false positive and false negative error rates, while a binomial-based maximum likelihood method gives information on error rates of individual bait proteins. This information is captured and published on a web interface (see companion abstract: "The Microbial Interactome Database: An Online System for Identifying Interactions Between Proteins of Microbial Species"). In addition, new decision tools are being developed to automate target selection with the goal of picking the protein targets that will best increase the resolution and breadth of the protein interaction networks.

37

The Microbial Interactome Database: An Online System for Identifying Interactions Between Proteins of Microbial Species

G.B. Hurst[1]* (hurstgb@ornl.gov), D.A. Pelletier[1], D.D. Schmoyer[1], M.B. Shah[1], W.H. McDonald[1], N.E. Baldwin[1], N.F. Samatova[1], A. Gorin[1], B.S. Hooker[2], V. Kery[2], W.R. Cannon[2], D.L. Auberry[2], K.J. Auberry[2], K.D. Victry[2], R. Saripalli[2], H.S. Wiley[2], S.J. Kennel[1], and **M.V. Buchanan**[1]

[1]Oak Ridge National Laboratory, Oak Ridge, TN and [2]Pacific Northwest National Laboratory, Richland, WA

The Center for Molecular and Cellular Systems (CMCS) is generating large amounts of protein-protein interaction data from microbial species through various "pipeline" protocols, as described in other abstracts. To make these data available to the scientific community, we are implementing a Microbial Interactome Database (MID) that will allow users to interact with the data at a variety of levels.

* Presenting author

The primary data source for the MID includes results from mass spectrometric identification of proteins interacting with affinity-tagged "target" or "bait" proteins, resulting from our endogenous and exogenous pipelines. We are currently acquiring these data for *R. palustris* and *S. oneidensis*. As of December 2005, we have performed affinity tagging, isolation, and mass spectrometric analysis for over 210 distinct target proteins ("baits") in *R. palustris,* and 53 proteins in *S. oneidensis*, with multiple replicate analyses in many cases. While rigorous and automated methods for distinguishing authentic interactors from non-specific interactions and background proteins are under development (as described in other abstracts from the CMCS), an empirical method based on filtering at several stages (individual mass spectra, comparison of estimates of interactor protein quantities with an appropriate average value over all comparable samples, frequency of observation, etc.) is currently in place. Interactors identified by this approach include some false positives (artifactual interactors), but also homologs to known complexes and potential novel interactions; experimental, literature-based, or informatic confirmation of these high-throughput results are described in other abstracts. This screen of four interaction subnetworks (DNA-directed RNA polymerase, DNA polymerase, ATP synthase and the degradosome) provides results that are consistent with homologous systems studied by other methods. For example, with α, β, and β' subunits as baits, we redundantly identify the core subunits of the RNA polymerase enzyme (α, β, β' and ω) as well as several sigma factors present only under certain conditions. For DNA polymerase, on the other hand, we have to date identified none of the expected interactors from experiments using β, τ/γ, δ, and χ subunits as baits in *R. palustris*; this result helps define the sensitivity of our current protocol, and suggests that larger culture sizes, more efficient isolation, or more sensitive detection will be required to study subnetworks representing complexes present at the level of ~10 copies per cell.

An internal project database tracks and summarizes data from all experiments, and provides access to CMCS investigators through a web interface. Publicly accessible views of selected pages of this website show a summary of the status of targeted proteins for the endogenous approach (http://maple.lsd.ornl.gov/cgi-bin/gtl_demo/public_target_status.cgi) and for the exogenous approach (http://maple.lsd.ornl.gov/cgi-bin/gtl_demo/public_ex_target_status.cgi,), and example interactor identifications for a selected subnetwork (http://maple.lsd.ornl.gov/gtl_demo/index.html).

We are currently designing a web-based resource to provide access to our results for the scientific community. Important components of this resource include presentation of our complete set of high-throughput "pipeline" results, comparison of putative interactions suggested by pipeline results with results from other databases of interacting protein and literature studies, and incorporation of results from confirmatory experiments such as live-cell imaging and surface plasmon resonance. Critical to the success of this resource will be an interactive filtering scheme to accommodate researchers with diverse scientific requirements, who will require different levels of stringency in the reliability of interactions reported, as well as different tools for extracting subsets of the data. Graphical representations (such as Cytoscape) of protein interaction subnetworks, incorporating various levels of experimental and comparative data, will be included.

Research sponsored by the Genomics:GTL program, Office of Biological and Environmental Research, U.S. Department of Energy, under contract No. DE-AC05-00OR22725 with Oak Ridge National Laboratory, managed and operated by UT-Battelle, LLC.

38

Investigation of Protein-Protein Interactions in the Metal-Reducing Bacterium *Desulfovibrio vulgaris*

Sara Gaucher[1,5]* (spgauch@sandia.gov), Masood Hadi[1,5], Swapnil Chhabra[1,5], Eric Alm[2,5], Grant Zane[3,5], Dominique Joyner[4,5], **Adam Arkin**[4,5], Terry Hazen[4,5], Judy Wall[3,5], and Anup Singh[1,5]

[1]Sandia National Laboratories, Livermore, CA; [2]Massachusetts Institute of Technology, Cambridge, MA; [3]University of Missouri, Columbia, MO; [4]Lawrence Berkeley National Laboratory, Berkeley, CA; and [5]Virtual Institute for Microbial Stress and Survival, http://vimss.lbl.gov

Desulfovibrio vulgaris is a sulfate reducing bacteria of interest due to its potential use in bioremediation as well as its economic impact in the petroleum industry (biocorrosion of pumping machinery). This sulfate reducing bacteria has been shown to reduce toxic metals (such as chromium and uranium) to insoluble species making them a good model system for understanding molecular machines involved in bioremediation of contaminated soils and ground water.

Protein complex isolation provides many challenges such as discrimination of non-specific vs. specific complexes and capture of transient partners. We have implemented two contrasting approaches for the isolation of protein complexes from *D. vulgaris*. The endogenous approach involves generating *D. vulgaris* cell lines that produce a "bait" protein of interest fused to an affinity tag (strep tag) that can be captured by chromatography. Lysate from these cells is passed over an avidin column, the bait protein with its associated proteins is captured and can be selectively eluted from contaminating proteins. In the exogenous approach, His-tagged bait proteins from *D. vulgaris* are expressed in *E. coli* cells. These bait proteins are purified to >90% purity and coupled to affinity beads which are then incubated with *D. vulgaris* lysate to capture interacting proteins. The advantage of the exogenous approach is that it is amenable to high throughput and automation. Further, the chance of capturing transient interactions may be increased by varying bait/lysate concentration to drive the equilibrium towards complex formation.

We computationally identified open reading frames (ORFs) that are involved in stress response (including oxygen, heat, pH and salt) based on homology to known stress related genes from other prokaryotic species and have used these ORFs as bait proteins to isolate their protein binding partners. Examples of such bait proteins studied include dnak, ClpX and CooX.. We have also selected proteins that are unique to this sulfate reducer ("signature" genes), expected to yield novel complexes related to sulfate/metal reduction. To validate our methods and address the challenge of non-specific binding, we have included some bait proteins whose interacting partners are well characterized in prokaryotic systems (*E. coli,* -rpoB and rpoC) and have validated our methods by isolating the binding partners of these targets.

In addition to direct complex isolation, we are also measuring protein expression in *D. vulgaris* under different stress conditions using Differential In-Gel Electrophoresis (DIGE) as a supplementary approach to identifying targets for protein-protein interactions studies. Potential binding partners may be gleaned from a list of proteins observed to be co-expressed.

* Presenting author

39

Protein Complex Analysis Project (PCAP): High Throughput Identification and Structural Characterization of Multi-Protein Complexes During Stress Response in *Desulfovibrio vulgaris*.

Project Overview

Dwayne Elias[3], Swapnil Chhabra[1], Hoi-Ying Holman[1], Jay Keasling[1,2], Aindrila Mukhopadhyay[1], Tamas Torok[1], Judy Wall[3], Terry C. Hazen[1], Ming Dong[1], Steven Hall[4], Bing K. Jap[1], Jian Jin[1], Susan Fisher[4], Peter J. Walian[1], H. Ewa Witkowska[4], **Mark D. Biggin**[1]* (mdbiggin@lbl.gov), Manfred Auer[1], Robert M. Glaeser[1], Jitendra Malik[2], Jonathan P. Remis[1], Dieter Typke[1], Kenneth H. Downing[1], Adam P. Arkin[1,2], Steven E. Brenner[1,2], Janet Jacobsen[2], and John-Marc Chandonia[1]

[1]Lawrence Berkeley National Laboratory, Berkeley, CA; [2]University of California, Berkeley, CA; [3]University of Missouri, Columbia, MO; and [4]University of California, San Francisco, CA

The Protein Complex Analysis Project (PCAP) has two major goals: **1**. to develop an integrated set of high throughput pipelines to identify and characterize multi-protein complexes in a microbe more swiftly and comprehensively than currently possible and **2.** to use these pipelines to elucidate and model the protein interaction networks regulating stress responses in *Desulfovibrio vulgaris* with the aim of understanding how this and similar microbes can be used in bioremediation of metal and radionuclides found in U.S. Department of Energy (DOE) contaminated sites.

PCAP builds on the established research and infrastructure of work by another Genomics:GTL project conducted by the Virtual Institute for Microbial Stress and Survival (VIMSS). VIMSS has developed *D. vulgaris* as a model for stress responses and have used gene expression profiling to define specific sets of proteins whose expression changes after application of a stressor. Proteins, however, do not act in isolation. They participate in intricate networks of protein / protein interactions that regulate cellular metabolism. To understand and model how these identified genes affect the organism, therefore, it is essential to establish not only the other proteins that they directly contact, but the full repertoire of protein / protein interactions within the cell. In addition, there may well be genes whose activity is changed in response to stress not by regulating their expression level but by altering the protein partners that they bind, by modifying their structures, or by changing their subcellular locations. There may also be differences in the way proteins within individual cells respond to stress that are not apparent in assays that examine the average change in a population of cells. Therefore, we propose to extend the VIMSS' analysis to characterize the polypeptide composition of as many multi-protein complexes in the cell as possible and determine their stoichiometries, their quaternary structures, and their locations in planktonic cells and in individual cells within biofilms. PCAP will characterize complexes in wild type cells grown under normal conditions and also examine how these complexes are affected in cells perturbed by stress or by mutation of key stress regulatory genes. These data will all be combined with those of the ongoing VIMSS project to understand, from a physical-chemical, control-theoretical, and evolutionary point of view, the role of multi-protein complexes in stress pathways involved in the biogeochemistry of soil microbes under a wide variety of conditions.

Essential to this endeavor will be the development of automated high throughput methods that are robust and allow comprehensive analysis of many protein complexes. Biochemical purification of endogenous complexes and identification by mass spectrometry will be coupled with *in vitro* and *in vivo* EM molecular imaging methods. Because no single method can isolate all complexes, we will

* Presenting author

develop two protein purification pipelines, one the current standard Tandem Affinity Purification approach, the other a novel tagless strategy. Specific variants of each of these will be developed for water soluble and membrane proteins. Our Bioinstrumentation group will develop highly parallel micro scale protein purification and protein sample preparation platforms, and mass spectrometry data analysis will be automated to allow the throughout required. The stoichiometries of the purified complexes will be determined and the quaternary structures of complexes larger than 250 Kd will be solved by single particle cryo EM. An innovative approach to discover weakly interacting protein partners will be investigated that combines rapid freezing of the whole contents of single bacteria, disrupted on the EM grid, and EM. EM tomography of whole cells and of sectioned, stained material will be used to detect complexes in cells and determine their structures at a resolution of 3 nm or better, which is sufficient to recognize known complexes by their characteristic sizes and shapes, to recognize the addition of labile components that may be lost during purification, and to identify instances of previously unknown, candidate complexes that are too labile to be purified by standard methods. New image analysis methods will be applied to speed determination of quaternary structures from EM data. Once key components in the interaction network are defined, to test and validate our pathway models, mutant strains not expressing these genes will be assayed for their ability to survive and respond to stress and for their capacity for bioreduction of DOE important metals and radionuclides.

40

Protein Complex Analysis Project (PCAP): High Throughput Identification and Structural Characterization of Multi-Protein Complexes During Stress Response in *Desulfovibrio vulgaris*

Microbiology Subproject

Terry C. Hazen[1]* (tchazen@lbl.gov), Dwayne Elias[3], Hoi-Ying Holman[1], Jay Keasling[1,2], Aindrila Mukhopadhyay[1], Swapnil Chhabra[1], Tamas Torok[1], Judy Wall[3], and **Mark D. Biggin**[1]

[1]Lawrence Berkeley National Laboratory, Berkeley, CA; [2]University of California, Berkeley, CA; and [3]University of Missouri, Columbia, MO

The Microbiology Subproject of the Protein Complex Analysis Project (PCAP) will provide the relevant field experience to suggest the best direction for fundamental, but DOE relevant research as it relates to bioremediation and natural attenuation of metals and radionuclides at DOE contaminated sites. We will build on techniques and facilities established by the Virtual Institute for Microbial Stress and Survival (VIMSS) for isolating, culturing, and characterizing *Desulfovibrio vulgaris*. The appropriate stressors for study will be identified and, using stress response pathway models from VIMSS, the relevance and feasibility for high throughput protein complex analyses will be assessed. We will also produce all of the genetically engineered strains for PCAP. Two types of strain will be constructed: strains expressing affinity tagged proteins and knock out mutation strains that eliminate expression of a specific gene. Over 300 strains expressing affinity tagged proteins will be produced every year for complex isolation and EM labeling experiments by the other Subprojects. A much smaller number of knockout mutation strains will be produced to determine the effect of eliminating expression of components of putative stress response protein complexes. Both types of engineered strains will be generated using a two-step procedure that first integrates and then cures much of a recombinant DNA from the endogenous chromosomal location of the

* Presenting author

target gene. For this we will develop new counter selective markers for *D. vulgaris*. This procedure will 1) allow multiple mutations to be introduced sequentially, 2) facilitate the construction of in-frame deletions, and 3) prevent polarity effects in operons. The Microbiology Subproject will provide high throughput phenotyping of all engineered strains to determine if any show phenotypic changes. We will test if the tagged proteins remain functional and that they do not significantly affect cell growth or behavior. The knockout mutations will be tested in a comprehensive set of conditions to determine their ability to respond to stress. High throughput optimization of culturing and harvesting of wild type cells and all engineered strains will be used to determine the optimal time points, best culture techniques, and best techniques for harvesting cultures using real-time analyses with synchrotron FTIR spectromicroscopy, and other methods. Finally, we will produce large quantities of cells under different conditions and harvesting techniques for optimal protein complex analyses. To insure the quality and reproducibility of all the biomass for protein complex analyses we will use rigorous QA/QC on all biomass production. We expect to do as many as 10,000 growth curves and 300 phenotype microarrays annually and be producing biomass for 500-1000 strains per year by end of the project. Each biomass production for each strain and each environmental condition will require anywhere from 0.1 – 100 L of culture, and we expect more than 2,000 liters of culture will be prepared and harvested every year. The Microbiology Subproject will optimize phenotyping and biomass production to enable the other Subprojects to complete the protein complex analyses at the highest throughput possible. Once the role of protein complexes has been established in the stress response pathway, we will verify the effect that the stress response has on reduction of metals and radionuclides relevant to DOE.

During the first 3 months, the Microbiology Subproject has supplied several sets of *D. vulgaris* biofilms for EM analysis, three 5 liter cultures of biomass for water-soluble protein complex purification studies, and a 120 liters culture for membrane protein complex purification. We have started using tandem affinity tagging of proteins using three distinct tags in order to purify the protein of interest for detailed characterization. The first uses a "*Strep*-tag" that inserts a streptavidin binding peptide for easy enrichment and we have currently constructed 16 such tags. To attain even higher protein enrichment, however, we are assessing the proven approach of a CTF (a.k.a. SPA) tag that includes a calmodulin binding protein (CBP), a protease (tobacco etch virus) and a 3 x FLAG sites for monoclonal antibody binding versus an "STF" tag that replaces CBP with a streptavidin binding peptide. At issue is the hypothesis that since the latter is only 8 amino acids compared to 125 for CBP, it should be less likely to interfere with localization/orientation of the protein within the cell. All three approaches are currently being assessed with DsrC (DVU2776), a protein in the dissimilatory sulfite reductase pathway that is essential for cell growth via sulfate respiration. Once we have confidently determined the best approach, we intend to tag at least 60 selected proteins in the coming year. Twenty of these proteins will also be tagged with a peptide including a tetracysteine motif that allows *in situ* EM imaging. Since the chemistry upon which this tag is based is thiol chemistry, we must first establish that the sulfide generated by these bacteria does not irrevocably interfere. FtsZ (DVU2499), a cell division protein, is the first candidate for testing the efficacy of this procedure. Within the time scope of this project, we intend to differentially tag >300 of the gene products in *D. vulgaris*. This information is expected lead to a more thorough understanding of not only the proteins involved in metal-reduction but also their protein-protein interactions and characterization of the complete pathway(s) for these activities.

41

Protein Complex Analysis Project (PCAP): High Throughput Identification and Structural Characterization of Multi-Protein Complexes During Stress Response in *Desulfovibrio vulgaris*

Multi-Protein Complex Purification and Identification by Mass Spectrometry

Ming Dong[1], Steven Hall[2], Bing K. Jap[1], Jian Jin[1], Susan Fisher[2], Peter J. Walian[1], H. Ewa Witkowska[2], and **Mark D. Biggin**[1]* (mdbiggin@lbl.gov)

[1]Lawrence Berkeley National Laboratory, Berkeley, CA and [2]University of California, San Francisco, CA

This Subproject of the Protein Complex Analysis Project (PCAP) proposes to test, automate, and use at high throughput multiple methods to purify *in vivo* protein complexes from *D. vulgaris*, identify their polypeptide constituents by mass spectrometry, and determine their stoichiometries. Our goal is to determine an optimum strategy that may include elements of each purification method. These methods will then be used as part of this project's broad goal of modeling stress responses relevant to the detoxification of metal and radionuclide contaminated sites.

Our first purification approach is a novel "tagless" method that fractionates the water soluble protein contents of a bacterium into a large number of fractions, and then identifies the polypeptide composition of a rational sampling of 10,000 – 20,000 of these fractions using MALDI mass spectrometry. Our second purification approach for water soluble proteins is to use and extend the proven Tandem Affinity Purification method (TAP), in which tagged versions of gene products are expressed in vivo and then used to purify the tagged protein together with any other endogenous interacting components. Our third and fourth approaches are specialized variants of the tagless and TAP methods that will be designed to capture membrane protein complexes. A major part of our effort will be the design and construction of automated instruments to speed the throughput of protein purification and sample preparation prior to mass spectrometry, and the development of rapid mass spectrometry data analysis algorithms.

Once established, we will use our optimized methods to catalog as thoroughly as practicable the repertoire of stable heteromeric complexes in wild type cells grown under normal conditions, as well as identify a number of larger homomeric complexes. We will then examine changes in the composition of protein complexes in cells with perturbed stress response pathways. Response pathways will be perturbed either by growing cells in the presence of stressors, including nitrite, sodium chloride, and oxygen, or by mutating cells to delete a component of a stress response pathway. Purified heteromeric and homomeric complexes larger than 250 kD will be provided to the EM Subproject to allow their structures to be determined and any stress induced changes in conformation to be detected. All of these data will be correlated by the Bioinformatics Subproject with computational models of stress response pathways that are currently being established by the Virtual Institute of Microbial Stress and Survival.

Our initial results to date are as follows. We have developed a partially optimized fractionation scheme for the tagless purification strategy and have used it to identify and purify several water soluble protein complexes from *D. vulgaris*. We have also isolated several putative membrane protein complexes following FOS-CHOLINE 12 detergent solubilization of membrane fractions and ion and molecular sieve chromatography. We have identified native gel electrophoresis as having great potential for high throughput, high resolution chromatographic separation of protein complexes

* Presenting author

and are now designing a prototype free flow electrophoresis device to allow use of this separation method within our pipeline. We are developing an iTRAQ based protocol that allows sufficiently accurate quantitation of relative protein levels by LC MALDI TOF/TOF mass spectrometry that co fractionation of polypeptides belonging to known protein complexes can be detected across a series of chromatographic fractions.

42

Protein Complex Analysis Project (PCAP): High Throughput Identification and Structural Characterization of Multi-Protein Complexes During Stress Response in *Desulfovibrio vulgaris*

Data Management and Bioinformatics Subproject

John-Marc Chandonia[1]* (JMChandonia@lbl.gov), Adam P. Arkin[1,2], Steven E. Brenner[1,2], Janet Jacobsen[2], Keith Keller[2], and **Mark D. Biggin[1]**

[1]Lawrence Berkeley National Laboratory, Berkeley, CA and [2]University of California, Berkeley, CA

The Data Management and Bioinformatics component of the Protein Complex Analysis Project (PCAP) has two major goals: **1.** to develop an information management infrastructure that is integrated with databases used by the Virtual Institute for Microbial Stress and Survival (VIMSS) project, and **2.** to analyze data produced by the other PCAP Subprojects together with other information from VIMSS to model stress responses relevant to the use of *D. vulgaris* and similar bacteria for bioremediation of metal and radionuclide contaminated sites.

The high-throughput experiments undertaken by the other PCAP Subprojects will require well engineered databases to capture and store the data in a consistent, structured format that renders it readily amenable to automated analyses. We will develop new infrastructure that is integrated with existing facilities currently deployed by the VIMSS project, in order to leverage existing tools for data analysis between the two projects. This infrastructure will be developed with an eye to scalability, in order to serve additional larger scale projects in the future. Processed data (i.e., complex composition and structures) will be integrated into web pages and other reports accessible through the VIMSS MicrobesOnline website, a public resource for comparative genomics research. Raw data will be made conveniently available to experimentalists for future reprocessing as necessary. We will assess the quality and consistency of all experimental data, as well as determine whether different experimental methods have more or less tendency to recover/detect intact complexes. We will compare our data to other public databases of protein complexes, pathways, and regulatory networks.

In the initial years of the PCAP project, we will prioritize proteins for tagging, TAP, and study by electron microscopy based on analysis of VIMSS data and other bioinformatic predictions. In our first round of prioritization, we have identified 10 *D. vulgaris* proteins as high-priority targets for tagging by the PCAP Microbiology Core, for experimental analysis by TAP/MS and electron microscopy.

All data we obtain on protein interactions will be analyzed in the context of the data currently stored in VIMSS. One of the primary goals of VIMSS is the creation of models of the stress and metal reduction pathways of environmental microbes. Protein-protein interaction data produced by PCAP will be used both to set the structure of these models as well as to parameterize them. Deletion data is particularly useful for inferring the structure of models. For example, if a deletion mutation leads

* Presenting author

to the absence of components from a complex, or the absence of entire complexes, this data may be used to infer details of cellular networks. We also hope to observe entire pathways being up- or down-regulated in response to stress. VIMSS models will be validated against the experimentally observed patterns of complex formation under different stress and deletion conditions.

43

Protein Complex Analysis Project (PCAP): High Throughput Identification and Structural Characterization of Multi-Protein Complexes During Stress Response in *Desulfovibrio vulgaris*

Imaging Multi-Protein Complexes by Electron Microscopy

Kenneth H. Downing[1]* (KHDowning@lbl.gov), Manfred Auer[1], Robert M. Glaeser[1], Jitendra Malik[2], Jonathan P. Remis[1], Dieter Typke[1], and **Mark D. Biggin[1]**

[1]Lawrence Berkeley National Laboratory, Berkeley, CA and [2]University of California, Berkeley, CA

The broad aim of this Subproject of PCAP is to demonstrate the feasibility of using electron microscopy for high-throughput structural characterization of multi-protein complexes in microbes of interest to DOE.

One goal of this work is to characterize the degree of structural homogeneity or diversity of the multi-protein complexes purified by PCAP and to determine the spatial arrangements of individual protein components within the quaternary structure of each such complex. It is already well-established that "single-particle" electron cryo-microscopy has unique capabilities for determining the overall quaternary structure of purified multi-subunit complexes whose molecular weight is greater than ~250 kDa. At a resolution of ~ 2 nm it is possible to locate the positions of individual proteins within such complexes and to then dock previously-determined atomic models of the identified proteins into the envelope of the density map. At resolutions better than 1 nm it is possible to further characterize conformational changes. We aim to increase the throughput of such structure determinations to the level that quaternary structures and docked atomic models are produced within 48 hours of purification of individual, structurally homogeneous complexes.

A second goal is to determine the spatial organization and relative locations of large multi-protein complexes within individual, intact microbes. It has quite recently been established that cryo-EM tomography can be used to produce clearly distinguishable images of larger multi-protein complexes (M_r > ~750 k) within suitably thin, intact cells. Since the cells are imaged in a nearly undisturbed condition, it is possible to count the number of such complexes in each cell as well as to characterize their spatial distribution and their association with other components of subcellular structure. This project will now seek to establish the extent to which it is possible to characterize significant changes in overall subcellular morphology that occur at the single-cell level in response to stress conditions, emphasizing the quantitative changes that occur in the temporal and spatial distributions of various multi-protein complexes. Cryo-tomography will thus be used to translate what is learned from purified multi-protein complexes, isolated from batch cultures, to the more complicated environment of intact cells.

A third goal is to determine whether whole-cell characterization by cryo-tomography can be further supplemented by electron microscopy of cell-envelope fractions and even the whole-cell contents of

* Presenting author

individual, lysed cells. Although both types of material no longer reflect the physiologically native conditions that exist within live microbes, both have a greater capability to gain information about complexes that are either too small to be recognized in tomograms of the molecularly crowded environment within thick specimens, or too labile to remain in tight association over the course of the protocols that must be used for biochemical purification. This approach may reveal an expanded population of multi-protein complexes, some of which may have been previously unknown and others of which may be composed of known, core complexes that still (i.e., immediately after lysis) retain one or more labile components.

Finally, plastic-section electron microscopy is used to translate as much as possible of this basic understanding to the more relevant physiological conditions, both stressed and unstressed, of planktonic and biofilm forms of microbes of interest. Although the morphological recognition of smaller protein complexes is less powerful for this type of electron microscopy, a compensating advantage is that this approach lends itself more easily to labeling – and thus localizing – genetically tagged proteins. Sectioning is also the only technique that can provide images of specimens that are too thick to image as whole-mount materials, while still retaining nanometer resolution. The ultimate goal in using plastic-section microscopy is thus to provide the most complete and accurate information possible about the status of multi-protein complexes, and to do so in a way that can then be used to improve mathematical modeling of cellular responses under the environmental conditions that require bioremediation.

Our initial experiments have been in the areas of single particle cryo EM and EM of biofilms of *D. vulgaris*.

Three separate samples of protein complexes have been evaluated by negative-stain, single-particle electron microscopy. Our pipeline calls for an initial evaluation of each such specimen in uranyl acetate, in neutralized phosphotungstic acid, and in ammonium molybdate, in order minimize misleading characterizations that inevitably occur due to unwanted stain-specimen interactions (e.g. spurious aggregation). One of the specimens appeared to be homogeneous and well dispersed, suitable for taking to the next step of single-particle cryo-EM. The other two specimens, however, appeared to have suffered from having been taken to too high a concentration prior to electron microscopy. From this experience we will adopt a modified protocol in which concentration of protein complexes to levels greater than are actually needed for electron microscopy is carefully avoided.

We have succeeded in growing biofilms of *D. vulgaris* in cellulose dialysis tubing and found the biofilms to cover almost the entire interior of the tube. We have successfully high-pressure frozen and freeze-substituted the biofilms. The sections examined by light microscopy reveal the expected overall biofilm architecture with channel-like areas that are devoid of bacteria or exopolysaccharide (EPS) material. Electron microscopic analysis of biofilm sections reveal loose packing of *D. vulgaris* within the biofilm EPS. Interestingly we found filamentous string-like metal precipitates near the *D. vulgaris*, which may point to structures not unlike the well-characterized *Shewanella* nanowires, which are known to be instrumental in extracellular metal reduction.

44

A Hybrid Cryo-TEM and Cryo-STEM Scheme for High Resolution *in vivo* and *in vitro* Protein Mapping

Huilin Li* (hli@bnl.gov), James Hainfeld, Guiqing Hu, Zhiqiang Chen, Kevin Ryan, and Minghui Hu

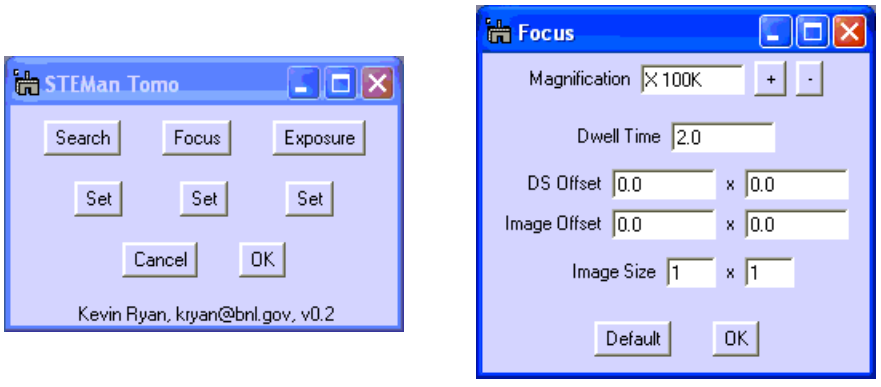
Brookhaven National Laboratory, Upton, NY

The overall structures of biological molecular assemblies and their locations inside a cell are keys to understanding their functions. Our *in situ* hybrid electron tomographic method, which takes advantage of ultra-structural visualization capability of the cryo-TEM and the heavy metal cluster label detection capability of the cryo-STEM, is capable of achieving simultaneously three-dimensional structural visualization and protein mapping. This method can be applied to map protein subunit positions inside a macromolecular assembly, or protein localization inside a microbial cell.

Our approach requires low dose electron imaging and tomographic capability in both TEM and STEM mode. Modern electron microscope comes with these capabilities only in TEM mode, but not in STEM mode. During the past year, we developed a Gatan Digital Micrograph Plug-In (we called it STEMan, see Figure 1) which was based on the Microsoft Visual Studio and Gatan's Software Developer's Kit (SDK). The Plug-In communicates with Jeol FasTEM communication interfaces for microscope setting and with Gatan Microscopy Suite 1.5 for electron beam scanning in STEM mode. With this interface, the damage to the specimen during search and focus can be avoided by offsetting scan and image positions. The process is conceptually similar to the low dose imaging in TEM mode.

Using above-described low dose imaging, we have examined the visibility of 2nm and 5nm unconjugated gold particles embedded in vitreous ice in both TEM and STEM modes. We found that 5 nm particles are visible in both TEM and STEM images, but STEM mode gives better signal to noise ratio. However, the 2nm gold particles in vitreous ice are not visible in TEM mode at low dose and low magnification (25KX) commonly used for tomography. The 2 nm particles are clearly visible in STEM mode at low dose and low magnification (25KX). From these in vitro studies, we expect that the 3 nm functionalized gold particles are detectable in cryo-STEM tomogram of a labeled protein complex either in solution or inside a cell.

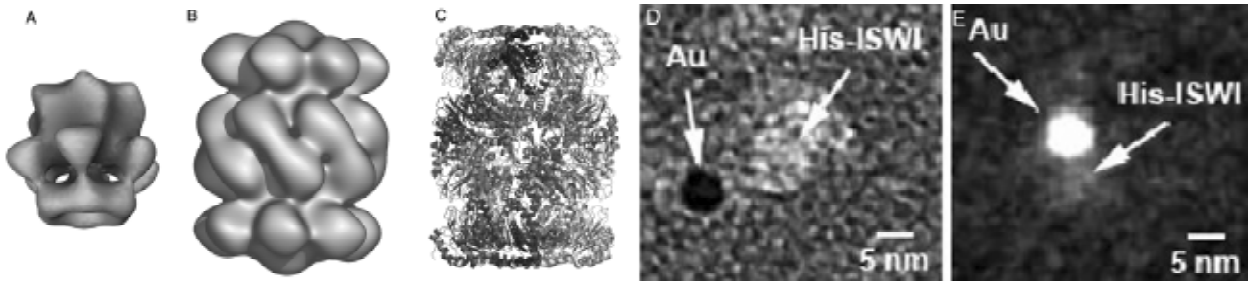
Figure 1. STEMan Plug-In graphic user interface. The main menu is on the left, the three top buttons change to the correct settings and then captures the image. The window to the right is displayed when the set button below Focus is pressed. This menu allows the magnification, exposure time, offset, and image size to be set. These values are saved and applied to the microscope when the corresponding image button is pushed.



* Presenting author

We have applied EM method to characterizing two microbial protein complexes - the Mycobacterial proteasome and proteasomal ATPase. Both complexes are implicated in resistance to macrophage killing. We found that the proteasomal ATPase is a hexameric structure with a large interior chamber and a relative flat bottom end (Figure 2A)[2]. Cryo-EM reveals a cluster density at the end of the proteasome (Figure 2B) [2,3], not resolved in crystallographic structure (Figure 2C)[3]. The density at the end blocks the entrance of substrate, strongly suggesting that the prokaryotic proteasomes are gated, contrasting to the general belief that the entrance is constitutively open. Our work demonstrates that cryo-EM reveals not only the architecture of microbial macromolecular assemblies, but also their slightly flexible domains that can be crucial to the functions.

Figure 2. Characterization of three macromolecular complexes by TEM and STEM. (A) Mpa, a 500 kDa proteasomal ATPase. (B) A 750 kDa proteasome with closed ends at the top and the bottom. (c) The absence of the end structure in the atomic structure. (D) TEM image of a 6x-histidine tagged protein (ISWI, 140 kDa) labeled with a 5 nm gold particle bearing the Nickel-NTA-group. (E) STEM image of the same sample as in (D).



We have synthesized 1.8 nm, 3 nm, and 5 nm functionalized gold particles. As an example of *in vitro* subunit mapping of multi-subunit complex, the His-tagged DNA remodeling complex is specifically labeled with the 5 nm gold particle with Ni-NTA group (Figure 2D and 2E). Further work will be focused on optimizing labeling condition and application of the technique to more microbial molecular machines both *in vitro* and *in vivo*.

References

1. "Characterization of a *Mycobacterium tuberculosis* proteasomal ATPase homologue," Darwin KH, Lin G, Chen ZQ, Li HL, Nathan, CF. *Molecular Microbiology*, 55, 561–571, (2005).
2. *"Mycobacterium tuberculosis* prcBA Genes Encode a Gated Proteasome with Broad Oligopeptide Specificity," Lin G, Hu GQ, Tsu C, Kunes Y, Li HL, Dick L, Li, P, Chen ZQ, Zwickl P, Weich N, Nathan C. *Molecular Microbiology*, in press.
3. "Structure of the *Mycobacterium tuberculosis* Proteasome and Mechanism of Inhibition by a Peptidyl Boronate," Hu GQ, Lin G, Wang M, Dick L, Xu R, Nathan C, Li HL. *Molecular Microbiology*, in press.

Acknowledgement: This research is supported by a grant from the Office of Biological and Environmental Research of the U.S. Department of Energy (KP1102010).

45

Optical Methods for Characterization of Expression Levels and Protein-Protein Interactions in *Shewanella oneidensis* MR-1

Natalie R. Gassman* (ngassman@chem.ucla.edu), Younggyu Kim, Sam Ho, Nam Ki Lee, Achilles Kapanidis, Xiangxu Kong, Gopal Iyer, and **Shimon Weiss**

University of California, Los Angeles, CA

Unraveling complex biological networks requires robust techniques capable of mapping protein-protein interactions and quantifying gene and protein expression levels. We have been developing a single optical technique, Alternating Laser Excitation (ALEX), which can integrate the analysis of protein-protein interactions and gene expression levels in an accurate, sensitive, and potentially high-throughput manner.

Previously, we reported on the capabilities of ALEX to expand single-pair Förster resonance energy transfer (spFRET) permitting ultra-sensitive analysis of biomolecular interactions by monitoring structure, dynamics, stoichiometries, local environment and molecular interactions. This is accomplished by obtaining D-excitation and A-excitation–based observables for *each single molecule* by rapidly alternating between a D-excitation and an A-excitation laser. This scheme probes directly both FRET donors and acceptors present in a single diffusing complex and recovers distinct emission signatures for all species involved in interactions by calculating two fluorescence ratios: the FRET efficiency *E,* a distance-based ratio which reports on conformational status of the species, and a new, distance-independent stoichiometry-based ratio, *S,* which reports on the association status of the species. Two-dimensional histograms of *E* and *S* allow virtual sorting of single molecules by conformation and association status (below). Using these ratios, we are now able to examine and characterize protein-protein interactions using both association and conformation information and quantify expression levels using the single-molecule sorting of association state.

We have continued to refine and extend the capabilities of the ALEX methodology, and through the addition of a third laser, 3-color alternating-laser excitation (3c-ALEX), we are now capable of measuring up to 3 intramolecular distances and complex interaction stoichiometries in solution. The 3c-ALEX system substantially extends 2-color ALEX by sorting the molecules in multi-dimensional probe-stoichiometry and FRET-efficiency histograms, and this multiplexing paves the way for advanced analysis of complex mixtures and biomolecular machinery at the single-molecule level. DNA model systems have been used to validate that 3c-ALEX permits FRET-independent analysis of three-component interactions; observation and sorting of singly-, doubly- and triply-labeled molecules present in the same solution; measurements of three intramolecular distances within single molecules, with no requirements for substantial FRET between the probes; and dissection of conformational heterogeneity with improved resolution compared to conventional single-molecule FRET.

To demonstrate the robust nature of these optical techniques and elucidate complex biological networks in a model genome, *Shewanella oneidensis* MR-1, we are applying the ALEX methodology to the regulatory mechanisms governing the transcription of genes in MR-1. Biomolecular interactions involved in signal transduction, formation the RNA polymerase-DNA transcription complex, and in activating or repressing transcription initiation can be monitored and analyzed with these techniques. We have successful reconstituted the transcription machinery from MR-1 and are beginning to examine the relevant protein-protein interactions that induce or repress transcription activation.

* Presenting author

In parallel with our efforts to examine protein-protein interactions in the transcription machinery of MR-1, we are also utilizing the ALEX technique for gene expression analysis. We have utilized DNA model systems to demonstrate the detection and quantification and are currently focusing our efforts at the mRNA level. Our progress in both of these areas will be reported.

46

Protein Interaction Reporter Studies on *Shewanella oneidensis* MR-1

James E. Bruce[1]* (james_bruce@wsu.edu), Xiaoting Tang[1], Wei Yi[1], Gerhard Munske[1], Devi Adhikari[1], Saiful Chowdhury[1], Gordon A. Anderson[2], and Nikola Tolic[2]

[1]Washington State University, Pullman, WA and [2]Pacific Northwest National Laboratory, Richland, WA

We have developed a unique chemical cross-linking system that employs novel compounds that we call "Protein Interaction Reporters" or PIRs that can help identify interactions among proteins in complex biological systems. We have previously applied this strategy to map interactions in a model noncovalent complex [1]. More recently, we have applied the PIR strategy to the microbial system, *Shewanella oneidensis* MR-1, to help elucidate protein interactions that facilitate novel electron transport mechanisms in this system. This report will describe the PIR strategy and highlight our initial PIR results with *S. oneidensis*.

The PIR strategy is based on the use of protein-reactive chemical functionalities that can covalently link interacting proteins in solution, complex mixtures, or within cells. This concept is common to a broad class of cross-linkers that have been exploited for protein structural analysis and limited protein interaction studies. Our PIR strategy combines the utility of chemical cross-linkers described above with mass spectrometry-cleavable features that can help differentiate multiple cross-linking reaction products and facilitate increased proteomics research. For example, the selective cleavage properties of PIR bonds allow release of intact peptides within the mass spectrometer. These peptides can then be studied independently with tandem MS and/or accurate mass analysis to produce data that allows protein identification. In doing so, interactions among proteins can be identified through multiple protein identities resultant from the peptides released from the PIRs. Additionally, the release of intact peptides masses allows differentiation of cross-linking reaction products such as dead-end, intra- and inter-molecular cross-linked species to be identified due to the mathematical relationships that exist between the precursor ions and observed masses that are released from the PIR-labeled species. Finally, the observation of the expected PIR reporter mass also facilitates internal calibration of tandem MS spectra, resulting in improved mass accuracy for tandem mass spectra and improved protein identification capabilities.

The PIR strategy was applied to intact, on-cell labeling studies with *S. oneidensis* following cell culture, harvesting and washing of cells. The current protocols developed for PIR studies utilize two stages of analysis to increase protein identification capabilities. Stage one is carried out with affinity capture of proteins from labeled cells, followed by digestion and tandem MS analysis. This results in a database of PIR labeled proteins. Sites of PIR incorporation are then determined through stage two analysis which involves digestion of all proteins after on-cell labeling, followed by affinity capture of the labeled peptides. These species are then subjected to multiplex LC-FTICR-MS experiments to measure both the intact PIR-labeled peptide masses, and the masses of the species released upon PIR activation. In summary, over 300 proteins were identified though the labeling experiments, many of which were found to be membrane or membrane associated proteins. Several

of these proteins are known to be critical for electron transport in this system. Additionally, many proteins were found to be plasma proteins and involved in protein synthesis and transport. Based on these initial results, we also carried out electron microscopy cell imaging experiments combined with PIR labeling to visualize sites of PIR incorporation on cell surfaces and interior to the cell plasma membrane. This presentation will highlight the sites of PIR localization, the proteins that were identified through the PIR approach, and discuss the potential for the future of these studies.

Reference

1. Tang, X., Munske, G.R., Siems, W.F. & Bruce, J.E. "Mass spectrometry identifiable cross-linking strategy for studying protein-protein interactions," *Anal Chem* **77**, 311-318 (2005).

47

Implementation of a Data Management and Analysis System In Support of Protein-Protein Interaction Studies of *Shewanella oneidensis* MR-1

Nikola Tolic[1]* (Nikola.Tolic@pnl.gov), Shaun O'Leary[1], Bryce Kaspar[1], Elena Peterson[1], Gunnar Skulason[2], Roger Crawford[2], Gordon Anderson[1], and **James Bruce**[2]

[1]Pacific Northwest National Laboratory, Richland, WA and [2]Washington State University, Pullman, WA

To address the data management and analysis needs of studying protein-protein interactions using cross-linkers and mass spectrometry we deployed the Data Management System (DMS). DMS is a part of the Proteomics Research Information Storage and Management System (PRISM) developed at Pacific Northwest National Laboratory (PNNL) for managing data at the Environmental Molecular Science Laboratory (EMSL) Proteomics Facility[1]. An FTICR mass spectrometer potentially produces hundreds of gigabytes of data daily so a key requirement for a data management system was to provide an affordable solution for data storage, archiving and backup. Since the Proteomics Facility in EMSL handles the same type of data on a much larger scale deploying DMS at Washington State University (WSU) to support their research efforts was a natural and logical selection. It also presented the opportunity to demonstrate DMS's flexibility and scalability for implementation in different environments.

Some key hurdles had to be overcome to deploy the custom built DMS solution to an outside facility. DMS was built to take advantage of PNNL's existing architecture and not necessarily for deployment elsewhere. Its security model is based on our internal network security and it makes extensive use of EMSL's multi-tera byte data archival system. WSU's infrastructure was configured to match PNNL's where DMS required the WSU system to connect to the EMSL archive securely. Because EMSL and PNNL provide a high level of data security collaborating with WSU electronically was a difficult technical and procedural task. The EMSL High Performance Computing & Networking Services group created new enabling technologies including new networks and new software to provide the secure access that was required.

As part of the protein-protein interactions research, the WSU team created a unique strategy they call Protein Interaction Reporter (PIR) which uses specially crafted cross-linker molecules[2] (see related abstract: "Protein Interaction Reporter Studies on *Shewanella Oneidensis* MR-1"). This approach requires development of software tools to analyze and transform the data to relevant information. Raw FTMS spectra are de-isotoped and interpreted using ICR-2LS[3] which was built

* Presenting author

in the EMSL and is integrated with DMS as an automated analysis. ICR-2LS is also used in combination with the commercial software tool called Mascot[4] for peptide and protein identification.

Another tool built specifically for this project is called *XLinks*. This is a set of tools built by at EMSL which employs custom macro functions incorporated into a Microsoft Excel™ template. *Xlinks* is used to extract and report putative cross-linked peptide products from the LC MS datasets by combining precursor and PIR fragmentation data to locate mathematical relationships inherent in data from PIR-labeled peptides. These tools allow automated assignment of PIR-labeled products from multiplexed LC-FTICR-MS datasets. Additional functions were developed and incorporated into *Xlinks* to allow quality control assessment and instrument performance validation on datasets of peptides of protein standards.

The deployment of this data management system to the WSU team has greatly enhanced their ability to continue research in peptide-peptide interactions. Because of the large amounts of data that is generated by the mass spectrometry manual analysis and interpretation becomes prohibitive. Additionally, automated analysis, such as the case with *Xlinks*, is not possible if data files are not stored in a manner that can be effectively accessed, processed, and analyzed. The DMS system at PNNL was an optimal solution to this process had never been deployed outside the PNNL campus. This report highlights the initial application of these critical proteomics tools to academic researchers, and demonstrates the benefits that can be gained by making these valuable National Lab Resources available to the rest of the scientific community. Also discussed in this presentation are the solutions that were devised to the technical and security hurdles that inhibited deployment of DMS to WSU researchers. These issues were critical to the successful implementation of DMS at WSU and the devised solutions are key to doing better science at WSU and other institutions that may implement DMS.

This research was supported by Office of Science (BER), U.S. Department of Energy, Grant No. DE-FG02-04ER63924.

References

1. Kiebel GR, Auberry K, Jaitly N, Clark DA, Monroe ME, Peterson ES, Tolic N, Anderson GA, Smith RD, "PRISM: A Data Management System for High-Throughput Proteomics," *Proteomics*, 2006. In Press.
2. Tang X, M.G., Siems WF, Bruce JE, "Mass spectrometry identifiable cross-linking strategy for studying protein-protein interactions," *Analytical Chemistry*, 2005. 1: p. 311-318.
3. ICR-2LS, http://ncrr.pnl.gov.
4. Perkins, D., D. Pappin, and et al., "Probability-based protein identification by searching sequence databases using mass spectrometry data," *Electrophoresis*, 1999. 20(18): p. 3551-3567.

48

Cell-Permeable Multiuse Affinity Probes (MAPs) and Their Application to Identify Environmentally Mediated Changes in RNA Polymerase and Metal Reducing Protein Complexes

M. Uljana Mayer, Baowei Chen, Haishi Cao, Seema Verma, Ting Wang, Ping Yan, Yijia Xiong, Liang Shi, and **Thomas C. Squier*** (thomas.squier@pnl.gov)

Pacific Northwest National Laboratory, Richland, WA

New generation cell-permeable multiuse affinity probes (MAPs) and complementary protein encoded tags have been developed and applied to identify the regulation of functional interactions between protein subunits in two key multiprotein complexes of *S. oneidensis*; i.e., RNA polymerase and metal reducing complex. Building upon a biarsenical scaffold, distinct reagents with differing colors have been constructed that have orthogonal sequence specificities, permitting their parallel application for high-throughput complex identification. Photoactivatable crosslinking moieties appended to MAPs provide a means to stabilize low-affinity binding proteins prior to cell lysis, providing the first robust means to mediate *in-vivo* crosslinking. An important advantage of this strategy is that a small and nonperturbing tag can be sequentially used to 1) measure the size of protein complexes and their binding affinities, 2) stabilize and isolate intact protein complexes for identification and structural analysis, 3) visualize the location and abundance of the protein complex within individual cells, and 4) the identification of protein function through the light-mediated inactivation of the tagged protein.

Specifically, we have shown:

1. To measure the size of protein complexes in the presence of other cellular proteins, we have taken advantage of the selectivity of MAPs to label tagged proteins, permitting the use of fluorescence correlation spectroscopy (FCS) to measure the size of the diffusing complex. Complementary measurements permit the titration of the protein complex, permitting the determination of the binding affinities of individual subunits.
2. Protein complexes were isolated following the immobilization of MAPs on glass supports, permitting the affinity isolation of tagged proteins and their binding partners. Release of the complex for analysis is facilitated by the mild reducing conditions associated with the release of the complex prior to mass spectrometric analysis. Stabilization of low affinity binding partners is facilitated by the application of newly developed cell permeable MAPs with photoactivatable moieties that permit *in vivo* cross-linking of binding partners.
3. Expression levels of tagged proteins can be directly visualized in live cells or following cell disruption and the separation of proteins on SDS-PAGE. Critical to this capability was the development of a new probe with increased polarity and charge that minimize nonspecific associations associated with high-background fluorescence common to commercially available dyes.
4. Targeted protein inactivation was demonstrated through light-induced protein inactivation, whereby the generation of singlet oxygen was shown to facilitate the formation of zero-length crosslinking and to selectively oxidize surface exposed methionines that are involved in the formation of protein-protein interfaces.

The combination of these capabilities resulting from a single genetic construct facilitates direct comparisons and provides the foundation for high-throughput measurements for a systems level understanding of cellular function.

* Presenting author

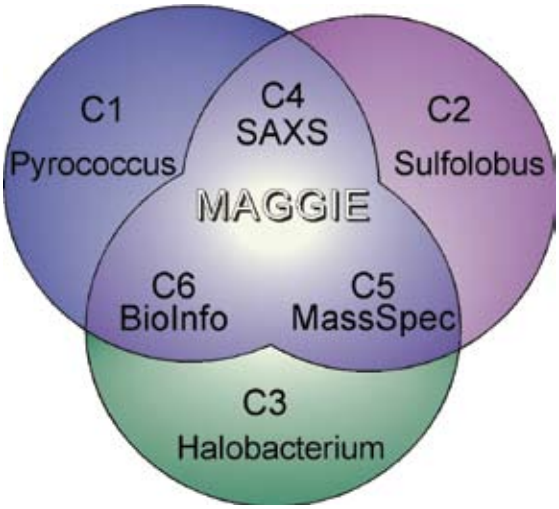
49

Molecular Assemblies, Genes, and Genomics Integrated Efficiently

John Tainer[1]* (jat@scripps.edu), Mike Adams[2], Steve Yannone[1], Nitin S. Baliga[3], Gary Siuzdak[4], and Steve Holbrook[1]

[1]Lawrence Berkeley National Laboratory, Berkeley, CA; [2]University of Georgia, Athens, GA; [3]Institute for Systems Biology, Seattle, WA; and [4]Scripps Research Institute, La Jolla, CA

MAGGIE is developing robust GTL technologies and comprehensive characterizations to efficiently couple gene sequences and genomic analyses with protein interactions and thereby elucidate functional relationships and pathways. The operational principle guiding MAGGIE overall objectives can be succinctly stated: protein functional relationships involve interaction mosaics that self-assemble from independent protein pieces that are tuned by modifications and metabolites. MAGGIE builds strong synergies among the Components to address long term and immediate GTL objectives by combining the advantages of specific microbial systems with those of advanced technologies. The objective for the 5-year MAGGIE GTL Program is therefore to comprehensively characterize the Protein Complexes (PCs) and Modified Proteins (MPs) underlying microbial cell biology. MAGGIE will address immediate GTL missions by accomplishing three specific goals: 1) provide a comprehensive, hierarchical map of prototypical microbial PCs and MPs by combining native biomass and tagged protein characterizations from hyperthermophiles (temperature-trapping otherwise reversible protein interactions) with comprehensive systems biology characterizations of a non-thermophilic model organism, 2) develop and apply advanced mass spectroscopy and SAXS technologies for high throughput characterizations of PCs and MPs, and 3) create and test powerful computational descriptions for protein functional interactions. An overall program goal is to help reduce the immense complexity of protein interactions to interpretable patterns though an interplay among experimental efforts of MAGGIE Program members in molecular biology, biochemistry, biophysics, mathematics, computational science, and informatics. In concert, MAGGIE investigators will characterize microbial metabolic modularity and provide the informed basis to design functional islands suitable to transform microbes for specific DOE missions. These efforts will furthermore test the degree to which metabolic and regulatory pathways can be treated as circuits in which PC and MP components can be swapped in and out to achieve GTL goals in microbial management.



50

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

Francis E. Jenney Jr.[1]* (fjenney@arches.uga.edu), Farris L. Poole II[1], Angeli Lal Menon[1], Rathinam Viswanathan[1], Greg Hura[2], John A. Tainer[2], Sunia Trauger[2], Gary Siuzdak[2], and **Michael W. W. Adams**[1]

[1]University of Georgia, Athens, GA and [2]Scripps Research Institute, La Jolla, CA

The genes that encode multiprotein complexes (PCs) or post-translationally modified proteins (MPs), such as those that contain metal cofactors, in any organism are largely unknown. We are using non-denaturing separation techniques coupled to mass spectrometry (MS) and metal (ICP-MS) analyses to identify PCs and MPs in the native biomass of a model hyperthermophilic organism of DOE interest, *Pyrococcus furiosus* (Pf), which grows optimally at 100 ˚C. By analyzing the native proteome at temperatures close to 100 ˚C below the optimum physiological temperature, we will trap reversible and dynamic complexes thereby enabling their identification and purification. Samples of the more abundant PCs and MPs obtained from native biomass are being used directly for structural characterization by Small Angle X-ray Scattering (SAXS), which provides information on overall mass, stoichiometry of subunits, radius of gyration, electron pair distances and maximum dimension. Recombinant versions of the less abundant PCs and MPs are being obtained by multiple gene expression systems designed using information from native biomass analyses and bioinformatic approaches. Robotic-based expression analyses are being used to assess the production of recombinant PCs and MPs, and these will be produced on a preparative scale for structural characterization. The recombinant aspect of the project utilizes the infrastructure developed for a previous structural genomics effort with Pf. This yielded the stable, recombinant forms of almost 400 Pf proteins. The non-recombinant aspect builds on extensive expertise in purifying and characterizing native multiprotein complexes from Pf using up to 1 kg of biomass as starting material. Results will be presented from complexes purified from Pf grown under at 95˚C using peptides as the primary carbon source.

* Presenting author

51

Systems Approach in a Multi-Organism Strategy to Understand Biomolecular Interactions in DOE-Relevant Organisms

Nitin S. Baliga[1] (nbaliga@systemsbiology.org) and John Tainer[2] (JATainer@lbl.gov)

[1]Institute for Systems Biology, Seattle, WA and [2]Lawrence Berkeley National Laboratory, Berkeley, CA

In this component of the MAGGIE project we will apply systems approaches to the study of molecular complexes. The important point to note is that this strategy will result in reduced costs by complementing through a versatile organism such as *Halobacterium* NRC-1 the lack of expensive systems biology tools for other organisms of the consortium. One of the many important contributions of this component will be the Gaggle. The Gaggle is a simple, open-source Java software environment that helps to solve the problem of software and database integration. Guided by the classic software engineering strategy of separation of concerns and a policy of semantic flexibility, it integrates existing popular programs and web resources into a user-friendly, easily-extended environment. We demonstrate that four simple data types (names, matrices, networks, and associative arrays) are sufficient to bring together diverse databases and software.