Proteomics and Metabolomics

111

Development and Application of New Technologies for Comprehensive and Quantitative High Throughput Microbial Proteomics

Richard D. Smith* (rds@pnl.gov), Mary S. Lipton, James K. Fredrickson, Matthew Monroe, Eric Livesay, Konstantinos Petritis, Joshua Adkins, Gordon A. Anderson, Kim Hixson, Ruihua Fang, Rui Zhao, Ronald J. Moore, and Yufeng Shen

Pacific Northwest National Laboratory, Richland, WA

With recent advances in whole genome sequencing for an increasing number of organisms, biological research is increasingly incorporating higher-level "systems" perspectives and approaches. Biology is transitioning from a largely qualitative descriptive science to a quantitative, ultimately predictive science. Key to supporting advances in microbial and other biological research at the heart of the DOE Genomics: GTL program is the ability to quantitatively measure the array of proteins (i.e., the proteome) in various biological systems under many different conditions. The challenges associated with making useful comprehensive proteomic measurements include identifying and quantifying large sets of proteins that have relative abundances spanning many orders of magnitude, which vary broadly in chemical and physical properties, have transient and low levels of modifications, and are subject to endogenous proteolytic processing. Ultimately, such measurements, and the resulting understandings of biochemical processes are expected to enable development of predictive computational models that could profoundly affect environmental clean-up and energy production by, for example, providing a more solid basis for mitigating the impacts of energy-production-related activities on the environment and human health.

In FY 2001, a project was initiated to develop quantitative and high throughput global proteomic measurement capabilities for microbial systems. The platform is based on a combination of advanced separations and mass spectrometric instrumentation and supporting computational infrastructure. The scope has included the development of an ultra-high pressure nano-scale capillary liquid chromatography platform combined with Fourier transform ion cyclotron resonance mass spectrometry and supporting data analysis and management capabilities. These developments provided the first high throughput mode "24/7" operation of such instrumentation, and resulted in its successful application to a set of microbial systems. The biological applications of this technology and associated activities are the subject of a separate, but interrelated project (J. K. Fredrickson, PI) involving a number of microbial systems (e.g. *Shewanella oneidensis* MR1, *Deinococcus radiodurans* R1*s*) in collaboration with leading experts on each organism. These studies have demonstrated the capability for automated high-confidence protein identifications, broad proteome coverage, and the capability for exploiting stable-isotope (e.g. ^{15}N) labeling methods to obtain high precision relative protein abundance measurements from microbial cultures.

A present emphasis of this project is the need for higher throughput proteomics measurements. A "prototype high throughput production" lab was established in FY 2002 was an early step in this direction. Operations within this lab are distinct from technology development efforts, both in laboratory space and staffing. This step was instituted in recognition of the different staff "mind sets"

required for success in these different areas, as well as to allow "periodic upgrades" of the technology platform in a manner that does not significantly impact its production operation. The result has been faster implementation of technology advances and more robust automation of technologies that improve overall effectiveness.

Our efforts currently in progress aim to:

- Significantly increase the overall data quality of global proteome measurements, and provide data that are quantitative and have statistically sound measures of quality.
- Increase overall data production by more than an order of magnitude in conjunction with the improved data quality.
- Provide the informatics tools and infrastructure required to support improved data quality and increased throughput and to efficiently manage, use, and disseminate large quantities of data generated by GTL "users."
- Develop the foundation for the further extension of proteomics measurements to enable more comprehensive coverage of protein modifications.

A significant challenge is the immense quantities of data that must be managed and effectively managed, analyzed, and communicated with associated measures of data quality in order to be useful. Thus, a key component of our program involves the development of the informatics tools necessary to make the data more broadly available and for extracting knowledge and new biological insights from large and complex data sets.

This research is supported by the Office of Biological and Environmental Research of the U.S. Department of Energy. Pacific Northwest National Laboratory is operated for the U.S. Department of Energy by Battelle Memorial Institute through Contract No. DE-AC06-76RLO 1830.

112

Characterization of *Rhodobacter sphaeroides* by High Resolution Proteomic Measurements

Mary S. Lipton*[1] (Mary.Lipton@pnl.gov), Timothy Donohue*[2] (tdonohue@bact.wisc.edu), Samuel Kaplan*[3] (Samuel.Kaplan@uth.tmc.edu), Stephen Callister[1], Matthew E. Monroe[1], Margie F. Romine[1], Ruihua Fang[1], Carrie D. Goddard[1], Nikola Tolic[1], Gordon A. Anderson[1], Richard D. Smith[1], Jim K. Fredrickson[1], Miguel Dominguez[2], Christine Tavano[2], Xiaihua Zeng[3], and Jung Hyeob Roh[3]

[1]Pacific Northwest National Laboratory, Richland, WA; [2]University of Wisconsin, Madison, WI; and [3]University of Texas Medical School, Houston, TX

Exploiting microbial function for purposes of bioremediation, energy production, carbon sequestration and other missions important to the DOE requires an in-depth and systems level understanding of the molecular components of the cell that confer its function. Inherent to developing this systems level understanding is the ability to acquire global quantitative measurements of the proteome (i.e. the proteins expressed in the cell). We have applied out state of the art proteomics technologies based upon high-resolution separations combined with Fourier transform ion cyclotron resonance mass spectrometry to obtain quantitative and high throughput global proteomic measurements of

* Presenting author

the photosynthetic bacterium *Rhodobacter sphaeroides.* Significant progress has been made addressing biological questions using high resolution proteomic measurements of cells, and fractions thereof, cultivated under varying conditions.

Rhodobacter sphaeroides 2.4.1 is α-3 purple nonsulfur eubacterium with an extensive metabolic repertoire. Under anaerobic conditions, it is able to grow by photosynthesis, respiration and fermentation. Aerobically it can grow by respiration as a chemoheterotroph. It can also be grown either photo- or chemo- lithotrophically on hydrogen and carbon dioxide. The organism can fix nitrogen under anaerobic conditions, and can use a wide diversity of terminal electron acceptors such as oxygen, metal oxides or oxyanions and an array of organic molecules as electron donors. When grown photosynthetically, it uses wavelengths of light in the near infra-red and contains a reaction center that is the ancestor of plant photosystem II. *R. sphaeroides* has been shown to possess two chromosomes, the larger of approximately 3.0 Mbp and the smaller of approximately 1.0 Mbp and 5 plasmids that together encode some 4600 gene products.

The initial mass tag database consisted of global proteomic preparations from the organism cultured under both steady state aerobic and photosynthetic conditions. However, important in the physiology of the organism is not just the global expression of proteins but also the localization of these proteins. For example, the transition of the organism between an aerobic to a photosynthetic state is accompanied by a synthesis of the photosynthetic membrane imbedded with the photosynthetic apparatus. It is therefore important to determine the localization of the proteins with in the organism to achieve a clear view of the physiology. To this end, cellular fractions of these organisms cultured under both highly aerobic conditions where photosynthetic membrane synthesis is repressed (30% O_2) and photosynthetic cell states (low, 3W/m^2, light intensity to maximize photosynthetic membrane synthesis) have been analyzed. Photosynthetic cells have been fractionated into 5 relatively discreet fractions (cytosol, periplasm, inner membrane, photosynthetic membrane and outer membrane) and the aerobic cells have been fractionated into 4 relatively discreet fractions (cytosol, periplasm, inner membrane, and outer membrane) in an effort to determine protein localization in the cell. We will report on the identification of ~XXXX total proteins from aerobic and photosynthetically-grown cells as well as the localization of proteins associated with assembly, function or control of the photosynthetic apparatus to individual subcellular fractions from steady-state photosynthetically grown cells.

The true understanding of the transition between the steady states will be achieved by a temporal study of the protein expression patterns in aerobically grown *R. sphaeroides* cells shifting to photosynthetic conditions. We have applied quantitative proteomics measurements to cells taken from a time course experiment of these cells transitioning between the two states. Preliminary studies are focused on the synthesis and deposition of the photosynthetic apparatus into the cells, however, through clustering analysis we will be able to identify other proteins that are important in this transition as well.

113

Quantitative Metalloproteomics

Patrick G. Grant* (pggrant@llnl.gov), Sharon Shields, Magnus Palmblad, and Graham Bench

Lawrence Livermore National Laboratory, Livermore, CA

Numerous bacteria have unusual enzymatic capabilities especially extreomophiles or bacteria that thrive in environments with extreme conditions (heat, acid, cold, etc…). Enzymatic activity commonly involves metal ions involved at the active site of the protein. To understand and utilize the enzymatic capability of these bacteria, enzymatic metalloproteins must be isolated, quantified, characterized, and identified as a function of the exposure to environmental cues(ions, pH, salt concentration). We have developed a nondestructive quantitative method to measure the amount of an isolated protein, the elements within that protein and identify the same protein sample with MALDI-TOF/MS by peptide mass fingerprinting.

Proteins are a critical class of biomolecule and the study of proteins, proteomics, has been enhanced with the sequencing of the genes that define proteins. However, the sequencing a species genome or even the analysis of the expression of these genes does not define the quantity of a protein within a biological partition or what is the function of the protein. The expression or production of proteins is estimated to range 8 orders of magnitude. Well-defined or isolated protein samples often contain little material. Purification of larger samples can be impractical (time or resources) or impossible (single patient), and more sensitive quantitation of small amounts of proteins is the practical solution.

It is estimated that one third of all proteins in eukaryotic species contain metal atoms with similar numbers expected in prokaryotic species, and many proteins contain molecular and elemental modifications for function and activation. Quantitation of these components within proteins has proceeded within individual protein studies instead of within groups of proteins or proteomes and is far from being completed. This neglects the relationships between the proteins and systematic changes. Accurate quantitation of proteins is also problematic because current methods utilize chemically-dependent spectroscopic analyses. The quantitative response from these analyses vary significantly even within a class of proteins. Quantitative responses are commonly linear only over limited ranges, requiring simultaneous calibrations using multiple internal or external standards. This is critical in proteomic studies of biological compartments, like a tissue, which can have up to 10000 different proteins expressed at once. We are utilizing nondestructive quantitation based on physical, rather than chemical, properties of the proteins to avoid these problems.

We utilize scanning transmission ion microscopy (STIM) to quantitate the sample, which is essentially dependent only on the amount of the analyte and is equally applicable to any isolated macromolecular sample. This technique quantitates the amount of the analyte by measuring the energy loss of a three million electron Volt (MeV) proton beam as it passes through a sample. This measurement method is independent of chemical structure and produces absolute analysis independent of standards. The sample rests on a thin, uniform substrate. The substrate allows further analysis of the same quantified sample for elemental content by particle induced X-ray emission (PIXE), identification of ligands by Time-of-flight secondary ion mass spectrometry (TOF-SIMS) or Hadamard Transform Time-of-flight mass spectrometry (HT-TOF-MS) and protein identification by Matrix Assisted Laser Desorption/ionization time-of-flight mass spectrometry (MALDI-TOF). Other surface based

* Presenting author

analytical techniques (RAMAN, UV/Vis, FTIR, etc…) and quantitation of isotopically labeled ligands with accelerator mass spectrometry (AMS) is also applicable to this same sample.

PIXE is an x-ray fluorescence technique that uses the same MeV proton beam to interrogate elemental composition within specimens. PIXE provides accurate quantitation, simultaneous multi-element detection for elements with atomic number greater than 12 and is capable of micron scale spatial resolution with 0.1 mg/kg elemental sensitivity.

We have developed methods for quantifying the mass of separated proteins at high femtomole levels, metal contents of the same protein sample to low femtomole sensitivity, identification to the femtomole level, and bound labeled ligands to the attomole level. Serial analysis of an individual sample allows the accurate determination of stoichiometric relationships of the protein to bound metals, many post-translational modifications, and bound ligands without the added error of duplicating samples for each analysis method. It is then critical that the protein that is quantitated and characterized must be identified which is possible since STIM/PIXE is nondestructive.

These measurements can be coupled to capillary or nanoscale liquid separation methods such as chromatography or capillary electrophoresis through the use of fraction collection onto our sample surface and volatile buffers. We have successfully deposited proteins from chromatography systems in spots as small as few tens of micrometers. This increases the range of mass quantitation down to a few nanograms with STIM. This project was supported by Laboratory Directed Research and Development funds.

114

New Technologies for Metabolomics

Jay D. Keasling* (jdkeasling@lbl.gov), Carolyn Bertozzi, Julie Leary, Michael Marletta, and David Wemmer

Lawrence Berkeley National Laboratory, Berkeley, CA

Microorganisms have evolved complex metabolic pathways that enable them to mobilize nutrients from their local environment and detoxify those substances that are detrimental to their survival. Metals and actinides, both of which are toxic to microorganisms and are frequent contaminants at a number of DOE sites, can be immobilized and therefore detoxified by precipitation with cellular metabolites or by reduction using cellular respiration, both of which are highly dependent on cellular metabolism. Improvements in metal/actinide precipitation or reduction require a thorough understanding of cellular metabolism to identify limitations in metabolic pathways. Since the locations of bottlenecks in metabolism may not be intuitively evident, it is important to have as complete a survey of cellular metabolism as possible. Unlike recent developments in transcript and protein profiling, there are no methods widely available to survey large numbers of cellular metabolites and their turnover rates simultaneously. The system-wide analysis of an organism's metabolite profile, also known as "metabolomics", is therefore an important goal for understanding how organisms respond to environmental stress and evolve to survive in new situations, in determining the fate of metals and actinides in the environment, and in engineering or stimulating microorganisms to immobilize these contaminants.

The goals of this project are to develop methods for profiling metabolites and metabolic fluxes in microorganisms and to develop strategies for perturbing metabolite levels and fluxes in order to study the influence of changes in metabolism on cellular function. We will focus our efforts on two microorganisms of interest to DOE, *Shewanella oneidensis* and *Geobacter metallireducens*, and the effect of various electron acceptors on growth and metabolism. Specifically, we will (1) develop new methods and use established methods to identify as many intracellular metabolites as possible and measure their levels in the presence of various electron acceptors; (2) develop new methods and use established methods to quantify fluxes through key metabolic pathways in the presence of various electron acceptors and in response to changes in electron acceptors; (3) perturb central metabolism by deleting key genes involved in respiration and control of metabolism or by the addition of polyamides to specifically inhibit expression of metabolic genes and then measure the effect on metabolite levels and fluxes using the methods developed above; and (4) integrate the metabolite and metabolic flux data with information from the annotated genome in order to better predict the effects environmental changes on metal and actinide reduction.

Recently, microorganisms have been explored for metal and actinide precipitation by secretion of cellular metabolites that will form strong complexes or by reduction of the metal/actinide. A complete survey of metabolism in organisms responsible for metal and actinide remediation, parallel to efforts currently underway to characterize the transcript and protein profiles in these microorganisms, would allow one to identify rate limiting steps and overcome bottlenecks that limit the rate of precipitation/reduction.

Not only will these methods be useful for bioremediation, they will also be useful for improving the conversion of plentiful renewable resources to fossil fuel replacements, a key DOE mission. For example, the conversion of cellulosic material to ethanol is limited by inefficient use of carbohydrates by the ethanol producer. Identification of limitations in cellulose metabolism and in products other than ethanol that are produced during carbohydrate oxidation could lead to more efficient organisms or routes for ethanol production – metabolomics is the key profile to identify these rate-limiting steps.

115

Characterization of Metal Reducing Microbial Systems by High Resolution Proteomic Measurements

Mary S. Lipton[1]* (Mary.Lipton@pnl.gov), Ruihua Fang[1], Dwayne A. Elias[1], Margie F. Romine[1], Alex Beliaev[1], Matthew E. Monroe[1], Kim K. Hixson[1], Yuri A. Gorby[1], Ljiljana Pasa-Tolic[1], Heather M. Mottaz[1], Gordon A. Anderson[1], Richard D. Smith[1], Jim K. Fredrickson[1], Derek Lovley[2], and Yanhuai R. Ding[2]

[1]Pacific Northwest National Laboratory, Richland, WA and [2]University of Massachusetts, Amherst, MA

Exploiting microbial function for purposes of bioremediation, energy production, carbon sequestration and other missions important to the U.S. Department of Energy (DOE) requires an in-depth and systems level understanding of the molecular components of the cell that confer its function. Inherent to developing this systems-level understanding is the ability to acquire global quantitative

* Presenting author

measurements of the proteome (i.e., the proteins expressed in the cell). We have obtained these types of measurements in a high throughput manner for the metal reducing bacteria *Shewanella oneidensis* and *Geobacter sulfurreducens* by application of our state of the art proteomics technologies based upon high-resolution separations combined with Fourier transform ion cyclotron resonance mass spectrometry. *S. oneidensis* MR-1, a Gram-negative, facultative anaerobe and respiratory generalist, is of interest to the DOE because it can oxidize organic matter by using metals such as Fe(III) or Mn(III,IV) as electron acceptors. This bacterium can also reduce soluble U(VI) to the insoluble U(IV) form, which prevents further U mobility in groundwater and subsequent contamination of down-gradient water resources. *Geobacter sulfurreducens* also is a dissimilatory metal-reducing bacterium that can reduce soluble U(VI) to insoluble U(IV). Such microbial reduction shows significant promise for *in situ* bioremediation of subsurface environments contaminated with U, Tc, and other toxic metals such as chromium.

In collaboration with the *Shewanella* Federation, we have characterized global cellular responses to changes in electron acceptors in *S. oneidensis* MR-1 cultures from a broad range of growth conditions, specifically the presence and absence of oxygen. This in-depth characterization requires not only a qualitative survey of the protein expression patterns, but also an understanding of how the levels of expression change with culture condition. To this end, we have applied quantitative proteomics approaches to characterize protein expression profiles in *Shewanella*. Relative changes in protein abundance were determined by both $^{14}N/^{15}N$ and absolute peak intensity. Each method of protein quantitation was applied to cells grown aerobically and anaerobically to determine the proteins important for electron transport within the cell. Proteins involved in the electron transport chain, as well as Fe(III) reduction were observed to increase in expression under anaerobic conditions, while proteins involved in pyruvate and malate synthesis were observed to decrease in expression under anaerobic conditions. All hypothetical and conserved hypothetical proteins are under evaluation for expression under these conditions, as well.

Other projects under the DOE Microbial Genome Program have already sequenced the *G. sulfurreducens* genome and initiated a functional genomics study to elucidate genes of unknown function in this organism. Proteomic efforts with this microorganism that complement this work have focused on creating a database of characteristic peptide mass and elution time tags, which serve as a unique 2D markers for subsequent peptide identifications. Initial global protein expression determinations have shown protein expression in most functional categories as assigned by TIGR. Thus far, quantitative analyses of protein expression patterns in *Geobacter* have been derived only by the absolute peak intensity method. Characterization of the multiple cytochrome proteins in the organism has shown interesting changes in these proteins when they are exposed to different electron acceptors. Additionally, extension of these studies to the clustering of the protein expression patterns is revealing interesting trends in proteins expression as a result of the variation in solubility among the electron acceptor.

The accuracy and precision of making these proteomic measurements is intricately linked to the analytical instrumentation, as well as to the efficiency of the sample processing methods. Advances in automation of sample processing will reduce variation among digested samples. Additionally, improved methods for quantitation and the application of increasingly sophisticated bioinformatics tools for data analysis will greatly improve the types and quality of the proteomic data available in the future.

116

Protein Complexes and Pathways

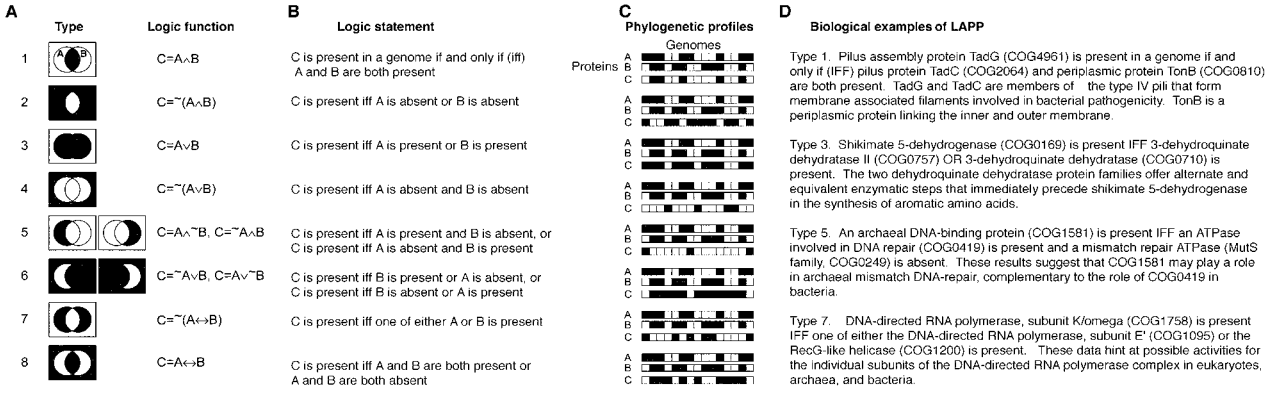
David Eisenberg* (david@mbiucla.edu), Peter Bowers, Michael Strong, Huiying Li, Lukasz Salwinski, Robert Riley, Richard Llwellyn, Einat Sprinzak, Debnath Pal, and Todd Yeates

University of California, Los Angeles, CA

Protein interactions control the life and death of cells, yet we are only beginning to appreciate the nature and complexity of their networks. We have taken several approaches towards mapping these networks. The first is the synthesis of information from fully sequenced genomes into knowledge about the network of functional interactions of proteins in cells. We analyze genomes using the Rosetta Stone, Phylogenetic Profile, Gene Neighbor, Operon methods to determine a genome-wide functional linkage map. This map is more readily interpreted when clustered, revealing groups of proteins participating in a variety of pathways and complexes. Parallel pathways and clusters are also revealed, in which different sets of enzymes operate on different substrates or with different cofactors.

These methods have been applied genome-wide to *Mycobacterium tuberculosis* and *R. Palustris*, as well as to more than 160 other genomes. Many results are available at: http://doe-mbi.ucla.edu/pronav The outcome is increased understanding of the network of interacting proteins, and enhanced knowledge of the contextual function of proteins. The information can be applied in structural genomics to find protein partners which can be co-expressed and co-crystallized to give structures of complexes.

In recent work, phylogenetic profiles have been extended by logical analysis of triplets of profiles, which reveal gene-encoded proteins that are involved in converging and parallel pathways, as well as linear metabolic pathways and complexes. The sorts of relationships uncovered are illustrated in the figure below. One example is that protein C may be present in a genome only if proteins A and B are both present. This type of logical analysis reveals many previously unidentified relationships in cellular networks because of branching and alternate pathways. It also facilitates assignment of cellular functions to uncharacterized proteins, and facilities mapping out of protein networks.



These inferred interactions can be compared to directly measured protein interactions, collected in the Database of Interacting Proteins: http://dip.doe-mbi.ucla.edu/.

* Presenting author

References

1. Visualization and interpretation of protein networks in *Mycobacterium tuberculosis* based on hierarchical clustering of genome-wide functional linkage maps. M. Strong, T.G. Graeber, M. Beeby, M. Pellegrini, M.J. Thompson, T.O. Yeates, & D. Eisenberg (2003). *Nucleic Acids Research*, **31**, 7099-7109 (2003).
2. Prolinks: a database of protein functional linkages derived from coevolution. P.M. Bowers, M. Pellegrini, M.J. Thompson, J. Fierro, T.O. Yeates, and D. Eisenberg (2004). *Genome Biology*, **5**:R35.
3. Use of Logic Relationships to Decipher Protein Network Organization. P.M. Bowers, S.J. Cokus, D. Eisenberg & T.O. Yeates, *Science*, **306**, 2246-2249 (2004).

117

Metabolomic Functional Analysis of Bacterial Genomes

Clifford J. Unkefer* (cju@lanl.gov)

Los Alamos National Laboratory, Los Alamos, NM

Parallel with the terms genome, transcriptome and proteome, the combined profile of cellular metabolites is the metabolome. Examining changes in the metabolome is a potentially powerful approach to assessing gene function and contribution to phenotype. Achieving the GTL goal of obtaining a complete understanding of cellular function will require an integrated experimental and computational analysis of genome, transcriptome, proteome as well as the metabolome. Moreover, metabolites and their concentrations are a product of cellular regulatory processes, and thus the metabolome provides a clear window into the functioning of the genome and proteome. The profile of metabolites also reflects the response of biological systems to genetic or environmental changes. In addition, metabolites are the effectors that regulate gene expression and enzyme activity. *The focus of this project is the elucidation of gene function by analysis of the metabolome.* We will carry out functional studies using stable isotope labeling and Mass or NMR spectral analysis of low-molecular weight metabolites. Like the proteome, metabolic flux and metabolite concentrations change with the physiological state of the cell. Because metabolite flux and concentration are correlated with the physiological state, they can be used to probe regulatory networks. In prokaryotic organisms, the combination of functional information derived from metabolic flux analysis with gene and protein expression data being developed in other laboratories will provide a powerful approach in identifying gene function and regulatory networks. Our pilot studies will build upon our capability, demonstrate the scientific value, and establish a facility for isotope-enhanced high throughput metabolome analysis of sequenced environmental microbes.

The power of metabolome analysis will be greatly enhanced by applying the combination of stable isotope labeling and mutations. Stable Isotope labeling and NMR/Mass spectral analysis of metabolites will be used to assign metabolic function in three ways. First, we will apply specifically labeled compounds to establish precursor product relationships, and test if putative pathways identified from analysis of the genome are operational. Next, we will develop the capability for functional genomic analysis using comparative metabolomics to reveal the phenotype of a set of so-called silent mutations. This method combines null mutants constructed from the genome sequence by allelic exchange with metabolomic analysis to elucidate the function of unknown ORF's. Finally, we will carry out a full metabolic flux analysis in steady state cultures. Flux analysis will provide input for a

stoichiometric model. Many of the advantages of isotope labeling for metabolomics in autotrophs and methylotrophs will be demonstrated throughout this proposal. Once demonstrated, this capability will be even more powerfully applied to heterotrophic organisms growing on complex substrates. These studies will lay the foundation to take similar labeling and metabolomic strategies into the environment to study microbial communities.

118

Dynameomics: Mass Annotation of Protein Dynamics through Molecular Dynamics Simulations of Fold-Space Representatives

David A. C. Beck* (dacb@u.washington.edu), Ryan Day, Kathryn A. Scott, R. Dustin Schaeffer, Robert E. Steward, Amanda L. Jonsson, Darwin O. V. Alonso, and Valerie Daggett

University of Washington, Seattle, WA

The Protein Data Bank (PDB) has been a tremendously useful repository of experimentally derived, static protein structures that have stimulated many important scientific discoveries. While the utility of static physical representations of proteins is not in doubt, as these molecules are fluid in vivo, there is a larger universe of knowledge to be tapped regarding the dynamics of proteins. Thus, we are constructing a complementary database comprised of molecular dynamics (MD) simulation [*1*] derived structures for representatives of all protein folds. We are calling this effort 'dynameomics.' For each fold (derived from consensus between SCOP, CATH, and DALI [*2*]) a representative protein is simulated in its native (i.e., biologically relevant) state and along its complete unfolding pathway by MD, the time-dependent integration of the classical equations of motion for molecular systems. There are approximately 1130 known non-redundant folds, of which we have simulated the first 30 that represent about 50% of known proteins. With the data resulting from our large database of MD simulations, we are data-mining for patterns and general features of transition, intermediate and denatured states to improve not only our understanding of protein dynamics but structure prediction, protein-protein and protein-ligand docking algorithms. Structure prediction remains one of the most elusive goals of protein chemistry. It is necessary to successfully predict native states of proteins, in order to translate the current deluge of genomic information into a form appropriate for functional identification of proteins from their primary sequence and rapid structure / dynamics based drug design. While these specific aims represent our immediate scientific goal for the dynameomics data, we are constructing a web site (http://www.dynameomics.org) for publication of the trajectories' coordinate data as well as in-depth analyses so that others may avail themselves of the resource and initiate areas of inquiry that we cannot even begin to anticipate.

References

1. Beck, D. A. C., and Daggett, V. (2004) Methods for Molecular Dynamics Simulations of Protein Folding / Unfolding in Solution, *Methods 34*, 112-120.
2. Day, R., Beck, D. A. C., Armen, R. S., and Daggett, V. (2003) A consensus view of fold space: Combining SCOP, CATH, and the Dali Domain Dictionary, *Protein Sci 12*, 2150-2160.

 * Presenting author